Homophily, influence and the decay of segregation in self-organizing networks

Adam Douglas Henry, Dieter Mitsche, Pawel Pralat

Homophily, influence and the decay of segregation in self-organizing networks

Adam Douglas Henry*
School of Government and Public Policy, University of Arizona, Tucson, AZ, USA
E-mail: adhenry@email.arizona.edu

Dieter Mitsche
Université de Nice Sophia-Antipolis, Laboratoire J-A Dieudonné
Parc Valrose, 06108 Nice cedex 02, France
‘ E-mail: dmitsche@unice.fr

Paweł Prałat†
Department of Mathematics, Ryerson University, Toronto, ON, Canada
E-mail: pralat@ryerson.ca

Abstract

We study the persistence of network segregation in networks characterized by the co-evolution of nodal attributes and link structures, in particular where individual nodes form linkages on the basis of similarity with other network nodes (homophily), and where nodal attributes diffuse across linkages, making connected nodes more similar over time (influence). A general mathematical model of these processes is used to examine the relative influence of homophily and influence in the maintenance and decay of network segregation in self-organizing networks. While prior work has shown that homophily is capable of producing strong network segregation when attributes are fixed, we show that adding even minute levels of influence is sufficient to overcome the tendency towards segregation even in the presence of relatively strong homophily processes. This result is proven mathematically for all large networks, and illustrated through a series of computational simulations that account for additional network evolution processes. This research contributes to a better theoretical understanding of the conditions under which network segregation and related phenomenon—such as community structure—may emerge, which has implications for the design of interventions that may promote more efficient network structures.

*A. D. Henry gratefully acknowledges support from the U.S. National Science Foundation (NSF) under grant #1124172, and the University of Arizonas Institute of the Environment.

†P. Prałat gratefully acknowledges support from NSERC and Ryerson University.

1 Introduction: Segregation and community structure in self-organizing networks

Many real-world networks exhibit network segregation, in that linkages are concentrated among nodes with shared or similar attributes [22, 32]. Related to network segregation, many networked systems also exhibit community structure, where nodes may be naturally partitioned into two or more groups of with many within-group linkages, but relatively few linkages across groups [26, 47, 51, 53]. These community boundaries often map onto a natural definition of groups based on clusters of nodal attributes; in this way, community structure is often a strong form of network segregation.[1]

Network segregation has important consequences for the efficient functioning of networks [36]. In some contexts it is advantageous for network nodes to have access, via relatively short paths, to other nodes with heterogeneous attributes or within distinct communities [8, 28]. For example, networks of information exchange among organizations and individuals involved in policy making often cluster around shared values or political positions, making it difficult for actors within these networks to reach negotiated agreements [56] or share information necessary to solve complex problems [35]. This is one explanation for the mismatch between scientific understanding and political action on complex policy issues such as climate change [19, 30]. Network segregation can also influence the ways in which technologies or behaviors diffuse in a network [3, 37], and may reinforce inequalities between groups of different socioeconomic status [36]. In some other cases, network segregation may be individually advantageous—for example, it may be useful for social actors to exploit fragmentations by occupying brokerage or bottleneck positions between communities [8]. Network segregation may even be globally advantageous if, for example, the presence of communities increases the stability and resilience of ecosystems [41] or provides a mechanism that shields "altruistic" or "cooperative" actors from the negative influence of noncooperative nodes [49]. In any case, better theories of the mechanisms that produce network segregation and community structure can inform the design of interventions to promote more efficient policy processes [55], increased social learning [30, 52], the emergence of cooperation [49], or any other process emerging from a broad array of network types [47].

Why does network segregation emerge? A large body of research has grown around methods for the efficient and reliable detection of communities in segregated networks [26, 53, 48], however relatively few studies have sought to develop theoretical explanations for why these network structures emerge and are maintained over time—but see [50, 32, 42, 43, 21, 12]. Research in this area often assumes some form of structural or choice-based homophily, in that linkages are more likely to form between similar nodes due to increased opportunities for interaction or systematic biases for homophilous links [6, 46].

The idea that nodes form connections with each other based on shared or similar traits is a strong and consistent empirical finding in the network science literature, particularly in the study of social networks. For example, research based on the Adolescent Health ("Add Health") dataset [29] shows a correlation between race and friendship among a sample of high school students, and that this homophily effect tends to be stronger for racial groups that comprise a larger proportion of the student population [14, 15]. An analysis of the same dataset also shows that genetic similarity—in addition to social and behavioral characteristics—also lead to the formation of friendship ties [5]. Among new university students in Germany, factors such as race, gender, and propensity for cooperative behavior were found to influence friendship choices [25]. In

[1] Community structure and network segregation are complementary but distinct concepts in network science. Community structure refers to a concentration of linkages among certain groups of nodes, whereas network segregation refers to a situation where the existence of links is correlated with the similarity of nodal attributes. It is possible to have networks that exhibit both segregation and community structure, but community structure is also possible without network segregation and network segregation is possible without community structure.

online social networks, homophily drives individuals to "befriend" others with similar tastes [45]. Within a large corporation, employees were shown to exhibit a preference for communication with others sharing their gender and basic job functions—at least when they had some flexibility to choose their communication partners [39]. A tendency toward homophily has also been observed for social aggregates. For example, organizations seeking to influence policy will tend to coordinate with other organziations that share their beliefs about policy problems [31, 33], and municipal governments tend to seek out collaborative ties with other governments on the basis of shared political ideology [24].

Within this very large literature on homophily, it should be noted that there is some variation in how the terms "homophily" and "segregation" are used. For our purposes, "homophily" refers to a dynamic process of tie formation or deletion, where the formation or deletion of ties occurs with probability proportional to the similarity or dissimilarity of network nodes. Homophily therefore refers to one of many possible factors that determine network structure. On the other hand, the term "segregation" is descriptive. Following [22], a segregated network is one where linkages are observed to be concentrated among nodes with similar traits—a characteristic that has sometimes been referred to as observed homophily [40]. Homphily is a natural explanation for segregation, and both empirical and theoretical research shows that even very weak forms of homophily can be a powerful force in the emergence of network segregation [7, 32, 40].

On the other hand, segregation may be produced through pathways other than homophily. In particular, many segregated networks emerge in the context of attributes that are malleable and shaped in part by the influences of other network nodes. Influence is therefore a potentially important mechanism of network evolution—under influence, a particular node's neighbors will exert a changing force on that node's attribute, making connected nodes more similar over time. Influence processes have been studied in many contexts including social learning and opinion formation [18, 23, 27], the diffusion of technological or policy innovations [57], and the adoption of culture or behaviors among human and non-human animals [16].

These mechanisms of network evolution—homophily and influence—are often studied in isolation, however these are often not independent processes. For example, research on social networks suggests that the degree to which two individuals exercise an influence on one another is proportional to the similarity of these individuals [10]. Moreover, in many contexts link structures and nodal attributes co-evolve with one another [44]. Studies such as [2] underscore the importance of considering both processes occurring in tandem; these authors show that homophily in a social network can change the efficiency of interventions meant to promote the spread of new technologies or behaviors.

Co-evolutionary models of homophily and influence are emerging [11, 17, 34, 45]. We advance work in this area by proposing a general mathematical model of network evolution where linkages are cut and formed through homophily, and where network nodes take on the attributes of others they are connected to. Our results show that even a minimal diffusion of attributes across linkages are sufficient to prevent the emergence of network segregation, even when individual nodes has a strong preference for segregation. We prove this result mathematically for all large network structures.

2 A mathematical model of network self-organization

The network evolution process analyzed here $\mathfrak{P}(G_0, \omega_0, \ell, q, r, K) = (G_t, \omega_t)_{t=0}^{\infty}$ is defined as a Markov chain and generates a stochastic sequence of networks (also referred to here as graphs). At each time step t, the process produces a graph $G_t = (V_t = V, E_t)$ consisting of $n = |V_t|$ nodes (also referred to as vertices) and $m = |E_t|$ links (also referred to as edges).[2] The total number of nodes and linkages does not change during

[2]It should be noted that the set of vertices does not change over time.

the network evolution process; thus, our attention focuses on the changing distribution of links and nodal attributes in the space over time, rather than on the growth of the network as studied in other lines of research [4, 20, 1, 13].

2.1 Measuring network segregation

During the network evolution process, individual nodes use their attributes to assess their similarity to or difference from other nodes, and nodal attributes also change as a result of influences exercised within network connections. Nodal attributes at time t are represented by the function $\omega_t : V \to [0,1]^r$, where $r \in \mathbb{N}$ represents the fixed dimensionality of attributes. The assumption that nodal attributes are multi-dimensional allows for more flexibility in modeling real-world processes where several unique attributes may exercise parallel influences on network structure. The similarity between any two network nodes is captured by a normalized measure of **attribute distance**, defined as $d : [0,1]^r \times [0,1]^r \to [0,1]$. For $x \in \mathbb{R}$, $\lceil x \rceil$ and $\lfloor x \rfloor$ are defined as the smallest integer being greater than or or equal to x and the largest integer being smaller than or equal to x, respectively. Following [32], attribute distances are converted into a discrete measure $d_K(\cdot,\cdot)$ by fixing $K \in \mathbb{N}$ and partitioning all distances into K bins such that

$$d_K(x,y) = \begin{cases} \frac{\lceil Kd(x,y) \rceil}{K} & \text{if} d(x,y) > 0, \\ \frac{1}{K} & \text{otherwise.} \end{cases}$$

The purpose of discretizing distances by introducing K bins is only to simplify calculations in this paper; our results do not depend on this approach. It should also be noted here that the actual distances $d(x,y)$ are assued to fall in the interval $[0,1]$, which ensures that any discretized distance d_K is no more than 1 and no less than $1/K$. Using this distance function, the case where $K = 1$ puts all attribute distances in the same bin, therefore "erasing" the effect of attributes on network structures.

The term **edge length** refers to the attribute distance of nodes that are connected by a particular network link—thus, actors with very similar attributes are connected by "short edges" whereas actors who are very different are connected by "long edges."

Networks comprised primarily of short edges are indicative of network segregation, which coupled with a multimodal distribution of attributes, also implies the existence of a community structure. Two additional variables help to track the emergence or suppression of network segregation. The **global center of mass** M_t is unique to each network G_t and is defined as the mean attribute across all nodes at time t,

$$M_t = \frac{1}{n} \sum_{v \in V_t} \omega_t(v).$$

The **local center of mass** $M_t(v)$ is unique to each network node v at a particular time step, and is defined as the mean attribute of that particular node's neighbors in the network:

$$M_t(v) = \frac{1}{|N_t(v)|} \sum_{u \in N_t(v)} \omega_t(u),$$

provided that $|N_t(v)| \geq 1$. $N_t(v) = \{u \in V_t : vu \in E_t\}$ denotes the set of v's neighbors at time t.

2.2 Dynamics of the evolutionary process: Homophily and influence

The evolutionary process begins at $t = 0$, with an **initial network** $G_0 = (V_0 = V, E_0)$ on $n = |V|$ vertices, $m = |E|$ edges, and with **initial vertex attributes** $\omega_0 : V \to [0,1]^r$. The network G_0 may have any structure and vertices might be distributed in any way in the attribute space. Time-step t, for $t \geq 1$, is defined to be the transition between G_{t-1} and G_t.

At time t, and independently of the history of the process prior to t, the network undergoes one of two processes: a **rewiring process**, where a random link is rewired via homophily, and an **influence process**, where attributes are transmitted through network ties. At each time step, influence occurs with probability $\ell \in (0,1)$; otherwise (that is, with probability $1-\ell$) the rewiring process is performed. The parameter ℓ is important because in most networks there will be far fewer nodes than linkages, and ℓ allows us to adjust the relative speed of influence and rewiring processes in terms of the proportion of all nodes and links that have been updated by these processes at least once. More information about the behavior of ℓ with respect to the relative speed of influence and rewiring is provided in the Supplemental Information.

When influence occurs, a random node $v \in V_{t-1}$ is chosen. Since influence is interpreted as a process of adopting the attributes of one's network neighbors, we assume that v's neighbors exert a force that pulls v towards its own local center of mass (that is, the average attribute of v's neighbors). The degree to which node v fully adopts the attributes of its neighbors may vary, and is captured by an additional parameter $q \in (0,1]$. If $q = 1$, then v jumps immediately to its local center of mass; for $q < 1$, on the other hand, v will move only part way. Thus, the parameter q represents the degree to which attributes are resistant to change. More precisely, if $N_{t-1}(v)$ is non-empty, then

$$\omega_t(v) = (1-q) \cdot \omega_{t-1}(v) + q \cdot M_{t-1}(v);$$

otherwise (that is, when $N_{t-1}(v) = \emptyset$), $\omega_t(v) = \omega_{t-1}(v)$. Attributes of other nodes do not change. A schematic of the influence process is depicted in the top panel of Figure 1.

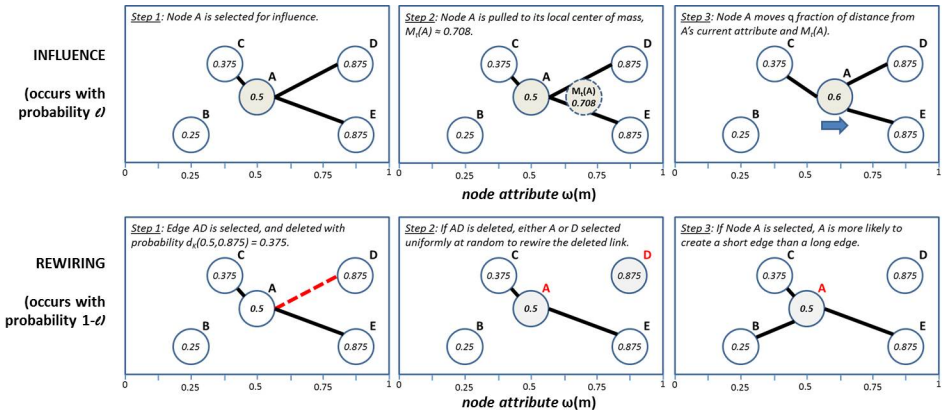

Figure 1: Network evolution process at time t. The figure depicts a schematic of the process for parameters $q = 0.5$ and $K = 8$. Nodal attributes are represented in the horizontal position of nodes. At time t, either a single node is selected for influence (probability ℓ) or a single edge is selected for rewiring (probability $1-\ell$). Under influence, the node selected for influence (A in the figure) is pulled to its local center of mass $M_t(A)$, or the average attribute of all neighbors. The attribute of A then shifts $q(M_t(A) - \omega(A))$, or from $\omega(A) = 0.5$ to $\omega(A) = 0.604$ in this example. Under rewiring, an edge (AD) is selected uniformly at random and is deleted with probability $d_K(\omega(A), \omega(D)) = 0.375$ (proportional to its edge length). A and D are then chosen with equal probability to rewire the link. The selected node (A in this case) assesses all possible edge lengths in the network and, under homophily, is more likely to create a link with a node having similar attributes than with a dissimilar node.

Homophily is modeled in the network rewiring process. When rewiring occurs, a random edge $uv \in E_{t-1}$ is chosen uniformly at random. The nodes u and v assess their attribute distance $d_K(\omega_{t-1}(u), \omega_{t-1}(v))$, and make a choice to rewire the linkage or to leave it in place. Due to homophily, nodes are assumed to prefer linkages with other nodes holding shared or similar attributes—thus, we assume that ties are rewired with probability proportional to their edge length (meaning that linkages between nodes with dissimilar attributes are more likely to be rewired). Specifically, the probability of tie deletion between u and v is $d_K(\omega_{t-1}(u), \omega_{t-1}(v))$.[3] When a tie is deleted, it is then rewired by selecting one of its endpoints with equal probability (set $w := u$ or $w := v$ according to the choice made), removing uv, choosing a vertex $x \in V_{t-1}$ at random, and then finally adding wx. The vertex x is selected so that creating short edges is more probable than creating long ones. Formally, the vertex $x \in V_{t-1}$ is chosen with probability proportional to $1/d_K(\omega_{t-1}(w), \omega_{t-1}(x))$; that is, with probability equal to

$$\frac{1/d_K(\omega_{t-1}(w), \omega_{t-1}(x))}{\sum_{v \in V_{t-1}} 1/d_K(\omega_{t-1}(w), \omega_{t-1}(v))} \geq \frac{1}{\sum_{v \in V_{t-1}} K} = \frac{1}{Kn}.$$

In order to get an upper bound, note that

$$\frac{1/d_K(\omega_{t-1}(w), \omega_{t-1}(x))}{\sum_{v \in V_{t-1}} 1/d_K(\omega_{t-1}(w), \omega_{t-1}(v))} \leq \frac{K}{\sum_{v \in V_{t-1}} 1} = \frac{K}{n}.$$

Thus, when a link is deleted, it is randomly assigned to one of the actors in the dyad, who then creates a new link with another network node chosen on the basis of attribute similarity.[4] This rewiring process is depicted in the bottom panel of Figure 1.

3 Results

As typical in random graph theory [38], all results are asymptotic, that is, for m tending to infinity, and thus for n tending to infinity as well. We say that an event holds *asymptotically almost surely* (*a.a.s.*) if it holds with probability tending to one as $m \to \infty$. Throughout, we will use the stronger notion of *with extreme probability* (*w.e.p.*) in favor of the more commonly used a.a.s., since it simplifies some of the proofs. We say that an event holds *w.e.p.* if it holds with probability at least $1 - \exp(\omega(m) \ln m)$ as $m \to 1$ ($\omega(m)$ is any function tending to infinity together with m). Thus, if we consider a polynomial number of events of which each holds w.e.p., then w.e.p. all events hold. For technical reasons, we will assume in our theorems that $\ln^7 n \ll m/n \ll n/\ln^3 n$.

This basic model framework allows for a rigorous analysis of systems characterized by varying levels of homophily and influence operating in tandem. Before moving to the technical arguments needed to analyze this model, it is useful to provide a brief synopsis of the main results.

3.1 Synopsis of findings

We show that even minute levels of influence are ultimately sufficient to cause vertex attributes to converge to a single point, even in the presence of strong homophily processes. This will happen more or less quickly

[3] Note that $d_K(\omega_{t-1}(u), \omega_{t-1}(v)) \in [1/K, 1]$ so the probability distribution is well-defined.

[4] This model allows loops and multiple edges, however there will typically be very few of these and their exclusion will not greatly affect the conclusions.

under different scenarios, and in some cases it is possible to predict with greater precision the process by which vertex attributes converge.

When the process begins at $t = 0$, vertex attributes may follow any distribution and the network G_0 may take on any initial structure. As the network undergoes subsequent rewirings, eventually each endpoint of each link has been rewired at least once—we call this time T. At this point, a sufficient amount of randomness has been introduced into the process, and we are able to predict the behavior of the model with high probability. We are able to establish an upper bound for T that is of order $m \log m$ (see Lemma 1).

After time T, it is useful to split the analysis of the network evolution process into two cases: first, where network rewirings do not take into account vertex attributes (the "zero homophily" case where $K = 1$, in which all edges are assigned the same distance, and thus independently of the underlying network, a rewiring is chosen uniformly at random), and second, where vertices cut and form ties on the basis of attribute similarity (the "homophily" case where $K \geq 2$; in which a rewiring is more likely to appear between close vertices). Analyzing the zero homophily case is useful as results from this more specific model may be applied to the analysis of models with either weak or strong homophily processes ($K \geq 2$ with K small or large, respectively).

In the zero homophily case, we can predict the specific process by which attributes converge. After time T, the local center of mass of each vertex is equal to the global center of mass in expectation, and in fact with high probability all local centers of mass are well concentrated around the global center of mass (see Lemma 2(i)). This means that the attributes of any given node's neighbor are similar to a random sample of vertex attributes drawn from the entire network. At this point in time, network structures will be similar to binomial random graphs. While vertex attributes have not necessarily converged at this point, these results imply that by time T any segregation seen in initial network conditions will have been destroyed by the influence processes without homophily supporting cross-group partitions. After at most $\tilde{T} = \Theta(\ell^{-1} n \ln n)$ more steps, all vertex attributes have then finally converged to the global center of mass. This result is given in Theorem 3.

In the homophily case, Lemma 2(ii) demonstrates that by time T, the local centers of mass are being pushed away from their boundaries towards the global center of mass. This means that the local centers of mass will typically be some fraction of the distance between that vertex's attribute and the global center of mass. We can specify an upper bound on this distance, which increases as the tendency towards homophily increases (that is, as K gets larger). Networks at this point exhibit network segregation as studied in [32] in that linkages in the network tend to be "short," that is, connecting vertices with similar attributes. If vertex attributes $\omega_t(v)$ maintain a multi-modal distribution at this point, then a clear community structure will emerge.

In both cases (that is, for any value of K), vertex attributes begin to converge after time T up to some time $T + \tilde{T}$. In the absence of homophily, all vertex attributes at time $T + \tilde{T}$ will be very close (within $o(\ln^{-2} n) = o(1)$) of the global center of mass (see Theorem 3).[5] The bigger n is, the closer all vertex attributes will be. This is because, for large n, the concentration results are stronger and the behavior of the whole network is even more predictable. An upper bound for $T + \tilde{T}$ is identified in Theorem 3. Moreover, we are able to predict precisely the process by which this convergence to the global center of mass takes place, in terms of how the average distance between each individual vertex and the global center of mass changes over time. The behavior of this random variable is predicted using the differential equation method and is specified in Theorem 12.

[5]We use the following standard asymptotic notation: $f(n) = o(g(n))$, if $\lim_{n\to\infty} f(n)/g(n) = 0$, and $f(n) = O(g(n))$ if there exists some n_0 so that for all $n \geq n_0$, $f(n) \leq Cg(n)$ for some absolute constant $C > 0$. Similarly, if $f(n) = O(g(n))$, then $g(n) = \Omega(f(n))$, and if both $f(n) = O(g(n))$ and $g(n) = O(f(n))$ hold, then $f(n) = \Theta(g(n))$.

In the presence of homophily, we determine an upper bound for the time when all vertex attributes are within $1/K$ of each other (see Theorem 5). At this point, the attributes of nodes are within a small enough interval that the distinctions between vertex attributes begin to disappear, meaning that homophily no longer exercises a noticeable influence on the network structure. From this point on, findings from the $K = 1$ case may be applied to show that, after another T time steps, the local centers of mass are well concentrated around the global center of mass as shown in Lemma 2(i).

The following sections analyze the model in more depth by asking (and answering!) a series of questions about the model's behavior with some constraints placed on parameters. Ultimately, these answers allow us to make precise predictions about how the model behaves without any parameter constraints. Much of the mathematical logic underlying these results is contained in the proofs, which are included in the Supplemental Information for interested readers.

3.2 How long does it take to rewire each edge endpoint at least once?

Each time a rewiring occurs, an edge is selected uniformly at random (that is, a given edge is selected with probability $1/m$), and then one of the endpoints of this edge is selected uniformly at random (that is, with probability $1/2$). So in each rewiring step, one of the $2m$ edge endpoints is selected uniformly at random and is associated with a random vertex. Although vertices that are more similar to the selected endpoint have a higher probability of being chosen, a given vertex is always selected with probability at least $1/(Kn)$. These steps occur independently of the history of the process.

Based on the following lemma, we know that the time at which each edge endpoint has been rewired at least once is no later than $\hat{T}$, which is of order $m \ln m$. A proof of Lemma 1 is provided in the Supplemental Information.

Lemma 1. *Consider the evolutionary process $\mathfrak{P}(G_0, \omega_0, \ell, q, r, K) = (G_t, \omega_t)_{t=0}^{\infty}$. Let $\omega(m)$ be any function that tends to infinity as $m \to \infty$, and let T be the first time every edge has both endpoints rewired at least once. Then, a.a.s.*

$$T \leq \frac{2Km}{1-\ell}\left(\ln m + \omega(m)\right) = \Theta(m \ln m). \tag{1}$$

It should be noted that this upper bound is a function of both K and ℓ. This is because, in addition to linkages being rewired, vertex attributes are also being updated as a result of the influence process. If the only process occurring were the rewiring of links, then from the well-known coupon collector problem it would be sufficient to wait $(1 + o(1))2m \log m$ time steps to rewire all $2m$ edge endpoints.[6] But since some time steps are used for influence rather than rewiring, it is necessary to wait $(1 + o(1))(2m/(1-\ell)) \log m$ rounds to ensure enough endpoints are rewired. Finally, since edge endpoints are not selected uniformly at random (i.e., endpoints associated with long edges have a higher chance of being selected), it is necessary to wait even slightly longer (a multiplicative factor of K is sufficient) to make sure that the endpoints of short edges are rewired as well.

[6] Let $c \in \mathbb{R}$. From the coupon collector problem, it follows that after $2m(\ln(2m) + c)$ rewirings, with probability tending to $e^{-e^{-c}}$, every endpoint is rewired at least once. From this it follows that if $c \to \infty$, then after $2m(\ln(2m) + c)$ rewirings, a.a.s. every endpoint is rewired at least once.

3.3 Where are local centers of mass relative to the global center of mass?

After the rewiring of every edge endpoint at time T, we can begin to estimate the distance between the global center of mass at time t (M_t) and the average attribute of vertex v's neighbors (v's local center of mass, denoted $M_t(v)$).

Suppose first that $K = 1$. In this situation, attributes' differences are irrelevant and edges are rewired without any tendency towards homophily; that is, regardless of whether the selected edge is long (connecting vertices with very different attributes) or short (connecting vertices with very similar attributes). The random graph at time T is relatively easy to study. In particular, for a given edge $e \in E_0$ from the initial network G_0, $t \geq T$, and any two vertices x and y, the probability that e connects x and y at time t is $2/n^2$. Indeed, with probability $1/n^2$ the first endpoint was associated with x and the second one with y, at the last time they were rewired. The same property holds when x is swapped with y. It follows that

$$\begin{aligned}\mathbb{P}(xy \in E_t) &= 1 - \left(1 - \frac{2}{n^2}\right)^m = 1 - \exp\left(-\frac{2m}{n^2} + O\left(\frac{m}{n^4}\right)\right) \\ &= \frac{2m}{n^2} + O\left(\frac{m^2}{n^4}\right) = \frac{2m}{n^2}\left(1 + O\left(\frac{m}{n^2}\right)\right) \\ &= (1 + o(\ln^{-3} n))\frac{2m}{n^2}, \end{aligned} \tag{2}$$

provided that $m = o(n^2/\ln^3 n)$.

In this situation, the expected degree of vertices in the network is $(1 + o(1))2m/n$, and networks exhibit properties similar to those observed in binomial random graphs $G(n, p)$ with $p = 2m/n^2$. In particular, the local centers of mass will be very close to the global one, provided that the network is dense enough to ensure that a given vertex's neighbors are representative of the entire set of network nodes.

Suppose now that $K \geq 2$. In this case, network rewiring is driven by some degree of homophily where long edges (connecting dissimilar vertices) are terminated with higher probability than short edges, and where short edges are formed with higher probability than long edges. This process tends to generate segregated, "attribute-close" networks where most edges are short, connecting vertices with similar or shared attributes; see [32] for more details. Therefore, it is expected that the local center of mass of a vertex v lies between v and the global center of mass.

Indeed, we are able to specify more precisely the distance between the local centers of mass and the global center of mass, in both the $K = 1$ and $K \geq 2$ cases. Since each dimension may be treated independently, we assume without loss of generality that $r = 1$. These distances are specified in Lemma 2.

Lemma 2. *Consider the social process $\mathfrak{P}(G_0, \omega_0, \ell, q, 1, K) = (G_t, \omega_t)_{t=0}^{\infty}$. Suppose that $d = 2m/n$ is such that $\ln^7 n \ll d = o(n/\ln^3 n)$. Let $C > 0$. Then, w.e.p. for every $T \leq t \leq n^C$ and every $v \in V_t$, the following holds*

(i) If $K = 1$, then

$$M_t(v) = (1 + o(\ln^{-3} n))M_t.$$

(ii) If $K \geq 2$, then

$$M_t(v) \geq (1 + o(1))K^{-6}M_t + o(\ln^{-5} n)$$

and

$$M_t(v) \leq 1 - (1 + o(1))K^{-6}(1 - M_t) + o(\ln^{-5} n).$$

At this point it is possible to have local centers of mass that are outliers, in the sense that vertices are closer to the global center of mass than their neighbors. In this case, local centers of mass can be pulled away from the global center because of the influence of their neighbors, but this does not substantially affect the global dynamics. However, once all vertex attributes are trapped in a particular interval, they will always remain within that interval. This observation—which we elaborate on in the following sections—tells us that local centers of mass that are outliers and close to the boundaries will tend to be pushed towards the global center of mass, causing the length of this interval in which all vertex attributes are "trapped" to shrink over time.

3.4 How quickly do vertex attributes converge in the absence of homophily?

We now consider the model's behavior in the absence of homophily—that is, when $K = 1$ and edge rewirings are independent of vertex attributes. These results may then be applied with some modifications to the unconstrained case where $K \geq 1$. For the sake of brevity, we state the final result for the $K = 1$ case here as Theorem 3. The Supplemental Information contains a formal proof of Theorem 3 as well as a full explanation of the logic of how this result is obtained through the analysis of a series of more simple models.

Theorem 3. *Consider the evolutionary process* $\mathfrak{P}(G_0, \omega_0, \ell, q, r, 1) = (G_t, \omega_t)_{t=0}^{\infty}$. *Suppose that* $d = 2m/n$ *is such that* $\ln^7 n \ll d = o(n/\ln^3 n)$ *and* $\ell \geq n^{-C}$ *for some* $C \geq 1$, *and recall the definition of* T *in (1). Then, a.a.s.*

$$\|\omega_{T+\tilde{T}}(v) - M_T\| = o(\ln^{-2} n) = o(1)$$

for every $v \in V_{T+\tilde{T}}$, *where* $\tilde{T} = \ell^{-1}(2+\varepsilon)n \ln n$ *for some* $\varepsilon > 0$.

Note that a.a.s.

$$T + \tilde{T} = \Theta(m \ln m) + \Theta(\ell^{-1} n \ln n) = O(n^{C+1} \ln n).$$

Theorem 3 says that, at time $T + \tilde{T}$ and for $K = 1$, the distance between each vertex's attribute and the global center of mass is $o(\ln^{-2} n)$; in words, the bigger n, the closer all vertex attributes are. The intuition behind this is that as the value of n grows, the process becomes more predictable (since the variance decreases by the law of large numbers). More precisely, all vertex attributes are in some interval (a, b) of length $o(\ln^{-2} n)$ and, once all attributes are in this interval, they will always remain there.

3.5 What is the process by which attributes converge in the absence of homophily?

Above, we demonstrated that vertex attributes will converge when $K = 1$. Additionally, it is possible to predict the convergence process between time T and $T + \tilde{T}$. This is done using the differential equation (DE) method [58], which allows us to show that a.a.s. the average distance between vertex attributes and the global center of mass at a given time period decreases in a predictable way. An introduction to the DE method is provided in the Supplemental Information.

Formally, let

$$W_t = \sum_{v \in V_t} |\omega_t(v) - M_t|.$$

The random variable W_t represents the overall deviations between the global center of mass and all individual local centers of mass at time t. Corollary 4 specifies how this random variable behaves over time.

Corollary 4. *Consider the evolutionary process* $\mathfrak{P}(G_0, \omega_0, \ell, q, r, 1) = (G_t, \omega_t)_{t=0}^{\infty}$. *Suppose that* $d = 2m/n$ *is such that* $\ln^7 n \ll d = o(n/\ln^3 n)$ *and* $\ell \geq n^{-C}$ *for some* $C \geq 1$. *Then, a.a.s.*

$$W_t = (1 + o(1))W_T \exp\left(-\frac{q\ell(t-T)}{n}\right)$$

for every t *such that* $T \leq t \leq T + t_f$, *where* T *is defined in (1) and*

$$t_f = \frac{n}{q\ell}\left(\ln\ln n - \ln\left(\frac{n}{W_T}\right)\right).$$

This result is a corollary to Theorem 12, which is explained in further detail (and proven) in the Supplemental Information.

3.6 Model behavior with homophily, $K \geq 2$

In a homophilous network, there is a bias for the termination of long edges and the creation of short edges. Suppose that all vertex attributes lie in the interval $[a, b]$ (as usual, we may treat all dimensions independently and so it is enough to focus on one dimension). It follows from Lemma 2(ii) that when vertex v is selected for influence at time $t \geq T$, it is moved away from at least one border of the attribute space (i.e., the value a or b). While the exact updated attribute of each vertex selected for influence cannot be predicted, it is known that the global center of mass is positioned far enough from one end of the interval such that all vertices selected for influence are pushed away from that end. Unfortunately, the global center of mass might be shifting during this process, and it is likely impossible to precisely predict how the global center of mass moves over time. This is in contrast to the result for the $K = 1$ case (see Theorem 3), which says that the global center of mass does not drift substantially between times T and $T + \tilde{T}$.

Fortunately, it is possible to show that after at most $3\ell^{-1}n\ln n$ steps, w.e.p. the maximum distance between any pair of nodes shrinks (in each dimension) by a multiplicative factor of at least $\varepsilon := q(13K^{12}n\ln n)^{-1}$. That is, if the distance is D at some time t, at time $t + 3\ell^{-1}n\ln n$ it is w.e.p. at most $D(1-\varepsilon)$. This is an upper bound only, and it is expected that the convergence process is in fact more rapid. A proof of Theorem 5 is provided in the Supplemental Information:

Theorem 5. *Consider the evolutionary process* $\mathfrak{P}(G_0, \omega_0, \ell, q, r, K) = (G_t, \omega_t)_{t=0}^{\infty}$. *Suppose that* $d = 2m/n$ *is such that* $\ln^7 n \ll d = o(n/\ln^3 n)$ *and* $\ell \geq n^{-C}$ *for some* $C \geq 1$. *Then, a.a.s. for every pair* $u, v \in V_t$ *we have*

$$\begin{aligned} |\omega_t(u) - \omega_t(v)| &\leq \sqrt{r}\left(1 - \frac{q}{13K^{12}n\ln n}\right)^{\lfloor \ell(t-T)/(3n\ln n)\rfloor} \\ &\leq \sqrt{r}\exp\left(-\frac{q\lfloor \ell(t-T)/(3n\ln n)\rfloor}{13K^{12}n\ln n}\right) \end{aligned}$$

for every t *such that* $T \leq t = O(\ell^{-1}n^2\ln^2 n)$ *(the random variable* T *is defined in (1)). In particular, if* $t = cq^{-1}\ell^{-1}n^2\ln^2 n$ *for some constant* $c > 0$, *then a.a.s. for every pair* $u, v \in V_t$ *we have*

$$|\omega_t(u) - \omega_t(v)| \leq (1 + o(1))\sqrt{r}\exp\left(-\frac{c}{39K^{12}}\right).$$

This result implies that at some time

$$\hat{T} \leq \left(\frac{1}{2}\ln r + \ln K + 1\right) 39K^{12}(q^{-1}\ell^{-1}n^2 \ln^2 n) = O(\ell^{-1}n^2 \ln^2 n)$$

a.a.s. all vertex attributes are within distance $1/K$ of each other. At this point, the attributes of nodes are within a small enough interval that the distinctions between vertex attributes begin to disappear, meaning that homophily no longer exercises a noticeable influence on the network structure. From this point on, findings from the $K = 1$ case may be applied to show that, after another T time steps, the local centers of mass are well concentrated around the global center of mass as shown in Lemma 2(i). With this result in hand, the behavior of the process can be predicted quite easily, and a.a.s. all the vertex attributes converge to a single point.

Corollary 6. *Consider the evolutionary process $\mathfrak{P}(G_0, \omega_0, \ell, q, r, K) = (G_t, \omega_t)_{t=0}^{\infty}$. Suppose that $d = 2m/n$ is such that $\ln^7 n \ll d = o(n/\ln^3 n)$ and $\ell \geq n^{-C}$ for some $C \geq 1$. Let $\hat{T}$ be the first time all pairs of vertices are within distance $1/K$ (or $\hat{T} = \infty$ if this never happens). Then, a.a.s.*

$$\hat{T} \leq \left(\frac{1}{2}\ln r + \ln K + 1\right) 39K^{12}(q^{-1}\ell^{-1}n^2 \ln^2 n) = O(\ell^{-1}n^2 \ln^2 n)$$

Moreover,

$$W_t = (1 + o(1))W_{\hat{T}+T} \exp\left(-\frac{q\ell(t - \hat{T} - T)}{n}\right)$$

for every t such that $\hat{T} + T \leq t \leq \hat{T} + T + t_f$, where T is defined in (1) and

$$t_f = \frac{n}{q\ell}\left(\ln\ln n - \ln\left(\frac{n}{W_{\hat{T}+T}}\right)\right).$$

In particular, at time $\hat{T} + T + t_f = O(\ell^{-1}n^2 \ln^2 n)$ the average distance between vertices is $o(1)$.

4 Examining dynamics through computational simulation

This analysis shows that, for all network structures, very small amounts of influence will lead to a convergence in attributes over time even when homophily (a polarizing force) is very strong. Thus, in theory, even small amounts of influence are sufficient to overcome even strong patterns of network segregation. The speed of convergence in attributes will of course depend on the relative strength of the homophily and influence processes. The preceding results allow us to predict with more precision the behavior of the system over time, in terms of how quickly the local centers of mass converge to the global center of mass for a given network. These dynamics are illustrated in Figure 2, which plots the maximum distance between local and global centers of mass over time, for 1) varying strengths of influence for a given level of homophily (left panel), and 2) varying levels of homophily for a given influence parameter (right panel).

The mathematical model developed here allows for a strong, generalized prediction about the behavior of networked systems over time. At the same time, the methods used here require limiting our attention to relatively simple network evolution processes, meaning that we face a fundamental trade-off between

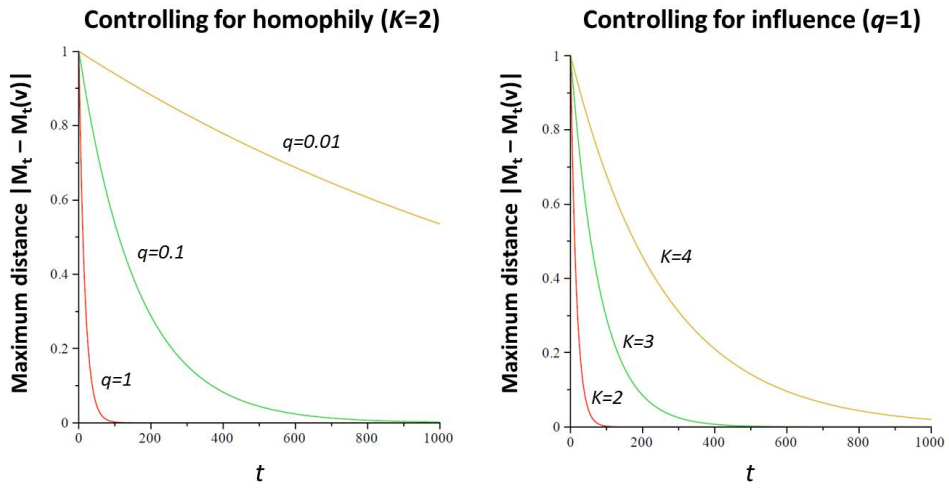

Figure 2: Upper bounds on attribute convergence speeds for varying levels of influence and homophily

model complexity and strength of predictions. Computational simulations are one way to introduce more complexity into these models, and also illustrate the evolutionary dynamics discussed in this paper. Thus, to complement the main mathematical model, we conducted a series of computational simulations of the influence and network rewiring process in small, random networks.[7]

4.1 Computational model setup

Our computational simulations begin with small, undirected random networks on 30 vertices and 44 edges (yielding a fixed network density of about 0.1). At initialization agents are assigned, uniformly and at random, an attribute value of 0 or 1. Links are then assigned with probability proportional to attribute similarity, so that initial networks exhibit high degrees of segregation and community structure (see, for example, the top panel of Figure 4). By starting with highly segregated initial network structures, we create conditions that are most favorable for the maintenance of network segregation over time.

With the initial networks in place, at each time step the system undergoes an influence process (where randomly selected vertices update their attributes based on network position), followed by an edge rewiring process (where randomly selected edges are deleted and reallocated to non-adjacent vertices).[8] When a vertex i is selected for influence, it moves a fraction of the distance between its own attribute and i's local center of mass (i.e., the average attribute of vertex i's neighbors). This fraction is governed by a model parameter $q \in [0, 1]$ as defined above.

Edge rewiring, on the other hand, progresses in a slightly different manner from the mathematical model presented in this paper. When an edge is selected for rewiring, it is randomly rewired by choosing a pair of

[7] Agent-based models were programmed in *R* [54]. Visualizations presented here made use of the *R* sna package [9].

[8] At each time step, 3 vertices are selected uniformly at random for influence and 9 edges are selected uniformly at random for rewiring. Among the selected vertices and edges, the influence and rewiring processes occur simultaneously. This setup was chosen to speed computation times.

non-adjacent vertices with probability proportional to the attractiveness of that pair. In order to account for more drivers of network formation, attractiveness of a potential link between two nodes i and j is allowed to be a function of three factors, including the similarity of nodal attributes (representing the tendency towards homophily), the number of paths of length 2 between i and j (representing the tendency towards transitivity, or so called *triadic closure*), and the degree of nodes i and j (representing the tendency towards forming linkages with high-degree actors, as in a preferential attachment model). In general, the attractiveness of a link between non-adjacent vertices i to j is assumed to be

$$a_{i \to j} = \frac{1}{d_K(i,j)} + p_{\text{closure}} T(i,j) + p_{\text{degree}} d(j),$$

where K is a simulation parameter representing the strength of the homophily effect, $d_K(i,j)$ is the discretized distance between vertices i and j (as described earlier), p_{closure} is a simulation parameter representing the strength of triadic closure, $T(i,j)$ is the number of paths of length 2 between vertices i and j, p_{degree} is a simulation parameter representing the strength of vertex degree, and $d(j)$ is the degree of j, that is, the number of vertices adjacent to j.

Thus, the attractiveness of two non-adjacent vertices depends on their distance, the degree of the incoming vertex, and the number of common neighbors of i and j. When an edge is selected for rewiring, the attractiveness scores of non-adjacent vertices are calculated. The resultant set of attractiveness scores is then converted into a probability distribution from which a particular dyad is selected—the edge selected for rewiring is then relocated to this randomly-chosen dyad.

The attractiveness of pairs of vertices based on these effects is illustrated in Figure 3. From the perspective of a particular vertex (labeled *Ego* in Figure 3), different potential linkages (non-adjacent vertices in the far left panel) have varying levels of attractiveness. Potential links are depicted as dashed red lines, and their attractiveness depends on the three simulation parameters *triadic closure*, *degree*, and *homophily* (K).[9] With higher triadic closure parameters, Ego will tend to view linkages that create triangles as more attractive, and therefore Ego will want to form a link with A in this schematic. With higher degree parameters, Ego will tend to view linkages with high degree vertices (e.g., vertex B) as more attractive. With higher homophily parameters, Ego will tend to view links as more attractive when the linkage reduces the difference between Ego's attribute and Ego's local center of mass (e.g., a link with vertex C).

4.2 Simulation results

Figure 4 displays one typical simulation run. On the left side of the figure are depictions of the network realized at each time step. In this diagram, the color of vertices indicate their starting attribute (0 or 1), and vertex sizes are proportional to the vertex's local center of mass and the network's global center of mass. The right side of the figure displays the distribution of the local centers of mass of each vertex, relative to the global center of mass (the global center of mass is indicated by the vertical reference line in the distribution plot). In this particular example, the local centers of mass collapse around the global center of mass after 65 time steps. This is also seen in the network diagrams, which show the decay of segregation around shared attributes early in the process to a well-mixed network where linkages are independent of starting attributes (node color) in the final time step.

Approximately 1,800 simulations were run with three model parameters—homophily, transitivity, and degree— assigned uniformly at random from $[0, 100]$ and q assigned uniformly at random from $[0.15, 1]$ at each

[9] The homophily parameter is analogous to the parameter K in the main mathematical model.

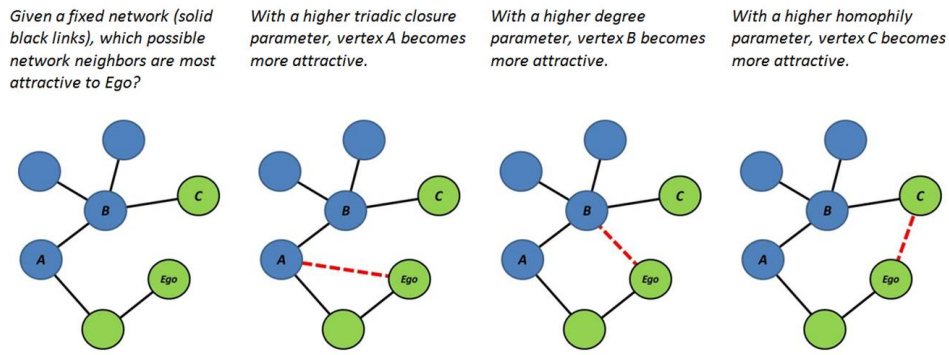

Figure 3: How simulation parameters change the relative attractiveness of pairs of vertices

simulation run. Consistent with our findings reported earlier, these simulations show that in all cases—even when additional network evolution dynamics are taken into account—vertex attributes always converge such that local centers of mass eventually collapse to the global center of mass. These dynamics are summarized in Figure 5. This figure plots the distribution of maximum distance between all local centers of mass and the global center of mass, for each simulation, and at different stages of the process. In order to visualize all simulations on a common scale, the distributions are represented at the initial time step in the far left, and subsequent distributions are for one-tenth of the final time step, two-tenths of the final time step, and so on. In the final time step, the maximum distance between the global and local centers of mass is no more than 0.05.

Figure 5 shows a large amount of variation in convergence patterns, however in aggregate we see a clear trend towards attribute convergence. But while attributes tend towards a single point, it is useful to examine how variation in the simulation parameters explain the speed of attribute convergence. Figure 6 summarizes these patterns by fitting an OLS linear regression model with time to convergence as the dependent variable, and simulation parameters as independent variables. Specifically, we model time to convergence as a linear combination of simulation parameters using the individual simulation run as the unit of analysis:

$$Y = b_0 + b_1 \cdot p_{\text{closure}} + b_2 \cdot p_{\text{degree}} + b_3 \cdot K + b_4 \cdot q + e,$$

where Y is a simulation's time to convergence, b_0, b_1, b_2, b_3, and b_4 are constant coefficients, e is an error term, and p_{closure}, p_{degree}, K, and q represent (respectively) the model parameters described above: the tendency towards triadic closure, the tendency to form ties with high-degree actors, homophily, and influence.

These results indicate that triadic closure and homophily have complementary effects on convergence speeds; both are positive, with roughly the same magnitude. This means that homophily and closure both tend to increase convergence times—that is, network segregation will take longer to disappear when tie formation is driven by homophily and triadic closure. This does make intuitive sense. If a network exhibits segregation, then triadic closure will tend to amplify the formation of ties among similar vertices because neighbors of neighbors will also tend to have similar attributes. On the other hand, introducing triadic closure does not prevent attributes from converging any more than homophily does.

Seeking out vertices with high degree appears to speed the process of convergence. This also makes intuitive sense, since this is a process that allows ties to be formed independently of attributes. So while

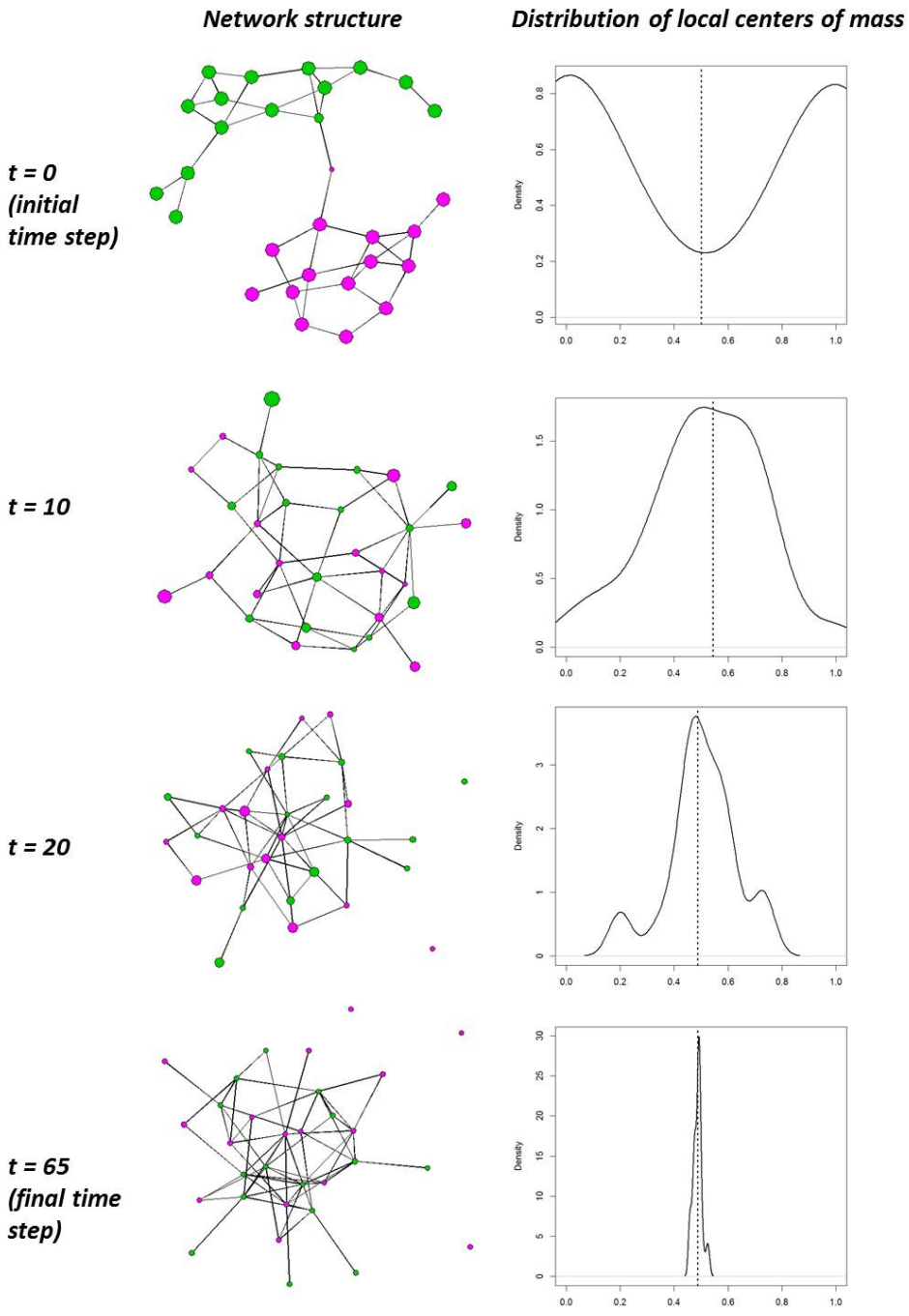

Figure 4: Evolution dynamics for a typical simulation run: the depicted simulation has parameters homophily (K) = 10; triadic closure (p_{closure}) = 30; degree (p_{degree}) = 30

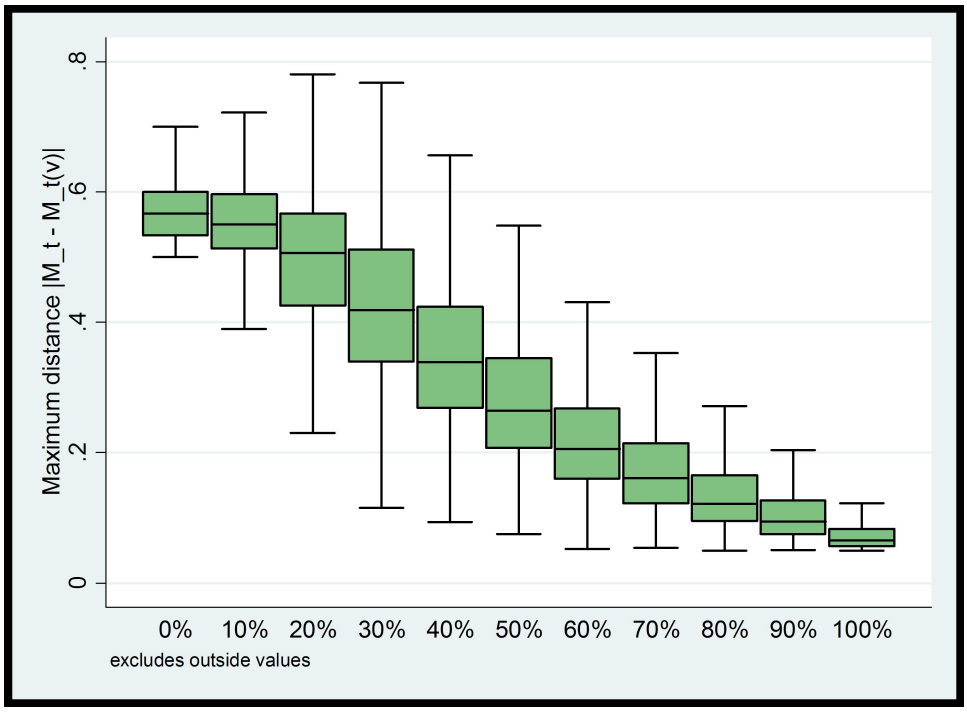

Figure 5: Aggregate simulation behavior over time

Attribute convergence times as a linear function of simulation parameters	
Variable	**DV = Time to attribute convergence**
Network evolution parameters	
Degree (p_{degree})	-0.135 *** (0.017)
Triadic closure ($p_{closure}$)	0.031 # (0.017)
Homophily (K)	0.046 ** (0.017)
Influence parameter	
Strength of influence (q)	-105.746 *** (1.975)
Constant	150.069
R^2	0.619
N (simulation runs)	1823
Note: Standard errors are given in parentheses. Stars indicate significance levels: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. Hash mark (#) indicates $p < 0.1$.	

Figure 6: Ordinary least squares regression model of effect of computational simulation parameters on time to attribute convergence

triadic closure will tend to create more short edges in the presence of network segregation, forming links on the basis of degree has nothing to do with attributes and provides a way for vertices to "escape" from homogenous network neighborhoods. The overall effect of degree-seeking is larger than both the homophily and triadic closure effects, meaning that degree-seeking behavior will tend to decay network segregation much faster than homophily or closure can maintain it.

5 Conclusion

This paper shows that influence processes are able to overcome the polarizing effect of homophily in creating segregated networks, even when strong forms of homophily are coupled with a relatively weak influence process. We prove mathematically that *asymptotically almost surely*, the size of the interval containing all vertex attributes tends to zero as t approaches infinity. In other words, the distinctions between vertices based on malleable attributes eventually disappear, which in turn prevents homophily from influencing emergent network structures.

Network segregation and related concepts, such as community structure, are well-studied phenomena in self-organizing networks. But why does network segregation emerge, and how is it maintained over time? Developing better models of these dynamics is important because they can suggest ways to manage the problems that arise from segregated networks—such as conflict or mistrust between social groups, or difficulties in searching for diverse sources of information.

While small amounts of homophily are sufficient to produce segregated networks when attributes are fixed, once malleability of attributes is introduced it becomes difficult to model the persistence of network segregation over time. We show mathematically that an important class of models, those where attributes move (potentially very slowly) towards one another, cannot maintain network segregation indefinitely. While our definition of influence represents one of many possibilities of how attributes change dynamically, these results are more powerful than those inferred from simulations that can only investigate a finite number of possible scenarios (and yet, the results are also supported by computational simulations that include more complex network evolution processes). With this research, we hope to provide a baseline for more investigations into the micro-level processes that generate and maintain the network segregation that is observed empirically across a wide range of self-organizing networks.

References

[1] Aiello, W., Bonato, A., Cooper, C., Janssen, J., and Prałat, P. (2009) A spatial web graph model with local influence regions, *Internet Mathematics* 5, pp. 175–196.

[2] Aral, S., Muchnik, L. and Sundararajan, A. (2013) Engineering Social Contagions: Optimal Network Seeding in the Presence of Homophily. *Network Science* 1, 125153.

[3] Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The Diffusion of Microfinance. Science, 341(6144), 12364981236498.

[4] Barabási, A.-L. and Albert, R. (1999) Emergence of Scaling in Random Networks, *Science*, vol. 286, no. 5439, pp. 509–512.

[5] Boardman, J. D., Domingue, B. W., and Fletcher, J. M. (2012). How Social and Genetic Factors Predict Friendship Networks. Proceedings of the National Academy of Sciences, 109(43), 1737717381.

[6] Boucher, V. (2015). Structural Homophily. International Economic Review, 56(1), 235264.

[7] Bramoullé, Y., Currarini, S., Jackson, M. O., Pin, P., and Rogers, B. W. (2012). Homophily and Long-Run Integration in Social Networks. Journal of Economic Theory, 147(5), 17541786.

[8] Burt, R. S. (2004) Structural Holes and Good Ideas. *Am J Sociol* 110, pp. 349–399.

[9] Butts, C. T. (2012) sna: Tools for Social Network Analysis. R package version 2.2-0. http://CRAN.R-project.org/package=sna

[10] Centola, D. (2011). An Experimental Study of Homophily in the Adoption of Health Behavior. Science, 334(6060), 12691272.

[11] Centola, D., Gonzalez-Avella, J. C., Eguiluz, V. M., and San Miguel, M. (2007). Homophily, Cultural Drift, and the Co-Evolution of Cultural Groups. Journal of Conflict Resolution, 51(6), 905929.

[12] Cooper, C., Frieze, A., and Prałat, P. (2014) Some typical properties of the Spatial Preferred Attachment model. *Internet Mathematics* 10, pp. 27–47.

[13] Cooper, C., Prałat, P. (2011) Scale-free graphs of increasing degree. *Random Structures and Algorithms* 38, pp. 396–421.

[14] Currarini, S., Jackson, M. O., and Pin, P. (2009). An Economic Model of Friendship: Homophily, Minorities and Segregation. Econometrica, 77(4), 10031045. http://doi.org/10.2139/ssrn.1021650

[15] Currarini, S., Jackson, M. O., and Pin, P. (2010). Identifying the Roles of Race-Based Choice and Chance in High School Friendship Network Formation. Proceedings of the National Academy of Sciences, 107(11), 48574861.

[16] Danchin, E., Giraldeau, L.-A., Valone, T. J. and Wagner, R. H. (2004) Public Information: From Nosy Neighbors to Cultural Evolution. *Science* 305, pp. 487–491.

[17] Dandekar, P., Goel, A., and Lee, D. T. (2013). Biased Assimilation, Homophily, and the Dynamics of Polarization. Proceedings of the National Academy of Sciences, 110(15), 57915796.

[18] DeGroot, M. H. (1974) Reaching a Consensus. *Journal of the American Statistical Association* 69, pp. 118–121.

[19] Dietz, T. (2013) Bringing Values and Deliberation to Science Communication. *Proceedings of the National Academy of Sciences* 110, pp. 14081–14087.

[20] D'Souza, R. M., Borgs, C., Chayes, J. T., Berger, N. and Kleinberg, R. D. (2007) Emergence of Tempered Preferential Attachment from Optimization. *Proceedings of the National Academy of Sciences* 104, pp. 6112–6117.

[21] Durrett, R. et al. (2012) Graph fission in an evolving voter model. *Proceedings of the National Academy of Sciences* 109, pp. 3682–3687.

[22] Freeman, L. C. (1978) Segregation in Social Networks. *Sociological Methods & Research* 6, 411429.

[23] Friedkin, N. E., and Johnsen, E. C. (2011) Social Influence Network Theory: A Sociological Examination of Small Group Dynamics, Cambridge University Press.

[24] Gerber, E. R., Henry, A. D., and Lubell, M. (2013). Political Homophily and Collaboration in Regional Planning Networks. American Journal of Political Science, 57(3), 598610.

[25] Girard, Y., Hett, F., and Schunk, D. (2015). How Individual Characteristics Shape the Structure of Social Networks. Journal of Economic Behavior & Organization, 115, 197216.

[26] Girvan, M. and Newman, M.E.J. (2002) Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences* 99, pp. 7821–7826.

[27] Golub, B., and Jackson, M. O. (2012). Does Homophily Predict Consensus Times? Testing a Model of Network Structure via a Dynamic Process. Review of Network Economics, 11(3).

[28] Granovetter, M. S. (1973) The Strength of Weak Ties. *The American Journal of Sociology* 78, pp. 1360–1380.

[29] Harris, K. M., Halpern, C. T., Whitsel, E., Hussey, J., Tabor, J., Entzel, P., and Udry, J. R. (2009). The National Longitudinal Study of Adolescent to Adult Health: Research Design. Retrieved from http://www.cpc.unc.edu/projects/addhealth/design

[30] Henry, A. D. (2009) The Challenge of Learning for Sustainability: A Prolegomenon to Theory. *Human Ecology Review* 16, pp. 131–140.

[31] Henry, A. D. (2011). Ideology, Power, and the Structure of Policy Networks. Policy Studies Journal, 39(3), 361383.

[32] Henry, A. D., Prałat, P. and Zhang, C.-Q. (2011) Emergence of Segregation in Evolving Social Networks. *Proceedings of the National Academy of Sciences* 108, pp. 8605–8610.

[33] Henry, A. D., Lubell, M., and McCoy, M. (2011). Belief Systems and Social Capital as Drivers of Policy Network Structure: The Case of California Regional Planning. Journal of Public Administration Research and Theory, 21(3), 419 444.

[34] Henry, A. D., and Prałat, P. (2013). Discovery of Nodal Attributes through a Rank-Based Model of Network Structure. Internet Mathematics, 9(1), 3357.

[35] Hong, L. and Page, S. E. (2004) Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers. *Proceedings of the National Academy of Sciences* 101, pp. 16385–16389.

[36] Jackson, M. O. (2014). Networks in the Understanding of Economic Behaviors. Journal of Economic Perspectives, 28(4), 322.

[37] Jackson, M. O. and López-Pintado, D. (2013) Diffusion and Contagion in Networks with Heterogeneous Agents and Homophily. *Network Science* 1, 4967.

[38] Janson, S., Łuczak, T., and Ruciński, A. (2000) Random Graphs. Wiley, New York.

[39] Kleinbaum, A. M., Stuart, T. E., and Tushman, M. L. (2013). Discretion Within Constraint: Homophily and Structure in a Formal Organization. Organization Science, 24(5), 13161336.

[40] Kossinets, G., and Watts, D. J. (2009). Origins of Homophily in an Evolving Social Network. American Journal of Sociology, 115(2), 405450.

[41] Krause, A.E., Frank, K. A., Mason, D. M., Ulanowicz, R. E. and Taylor, W. W. (2003) Compartments Revealed in Food-Web Structure. *Nature* 426, 282285.

[42] Kumpula, J., Onnela, J.-P., Saramäki, J., Kaski, K. and Kertész, J. (2007) Emergence of Communities in Weighted Networks. *Physical Review Letters* 99.

[43] Lambiotte, R., Ausloos, M. and Holyst, J. (2007) Majority Model on a Network with Communities. *Physical Review* E 75.

[44] Lazer, D. (2001) The Co-Evolution of Individual and Network. *The J. of Math. Sociology* 25, pp. 69–108.

[45] Lewis, K., Gonzalez, M., and Kaufman, J. (2012). Social Selection and Peer Influence in an Online Social Network. Proceedings of the National Academy of Sciences, 109(1), 6872.

[46] McPherson, M., Smith-Lovin, L. and Cook, J. M. (2001) Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27, pp. 415–444.

[47] Newman, M. E. J. (2003) The Structure and Function of Complex Networks. *SIAM Review* 45, pp. 167–256.

[48] Newman, M. E. J. (2006) Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences* 103, pp. 8577–8582.

[49] Nowak, M. A. (2006) Five Rules for the Evolution of Cooperation. *Science* 314, pp. 1560–1563.

[50] Palla, G., Barabási, A.-L. and Vicsek, T. (2007) Quantifying social group evolution. *Nature* 446, pp. 664–667.

[51] Palla, G., Derényi, I., Farkas, I. and Vicsek, T. (2005) Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. *Nature* 435, pp. 814–818.

[52] Parson, E. A. and Clark, W. C. (2005) in Barriers and Bridges to the Renewal of Ecosystems and Institutions, pp. 428–460, Columbia University Press.

[53] Porter, M. A., Onnela, J.-P. and Mucha, P. J. (2009) Communities in Networks. *Notices of the American Mathematical Society* 56, pp. 1082–1097.

[54] R Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[55] Sabatier, P. E. (2007) Theories of the Policy Process. Westview Press.

[56] Sabatier, P. E. and Jenkins-Smith, H. C. (1993) Policy Change and Learning: An Advocacy Coalition Approach. Westview Press.

[57] Valente, T. W. (1995) Network Models of the Diffusion of Innovations. Hampton Press.

[58] Wormald, N. (1999) The differential equation method for random graph processes and greedy algorithms. In: Lectures on Approximation and Randomized Algorithms, eds. M. Karoński and H. J. Prömel, PWN, Warsaw, pp. 73–155.

6 Supplemental Information

6.1 Introduction to the Differential Equation method

The general setting that is used in the DEs method [58] is a sequence of random processes indexed by n (which in our case is the number of objects in any set V_t). The aim is to find properties of the random process in the limit as $n \to \infty$. The conclusion we aim for is that variables defined on a random process are well concentrated, which informally means that with high probability they are very close to certain deterministic functions. These functions arise as the solution to a system of ordinary first-order differential equations. One of the important features of this approach is that the computation of the approximate behavior of processes is clearly separated from the proof that the approximation is correct.

To show that the random variables in a process usually approximate the solution of differential equations, we need to use large deviation inequalities. These inequalities are often used to give an upper bound on the probability that a random variable deviates very far from its expected value. In a typical situation with a random process, the aim is to show that the random variable Y_t of interest is sharply concentrated. In fact,

$$Y_t - Y_0 = \sum_{i=1}^{t} (Y_i - Y_{i-1}).$$

If the differences $Y_i - Y_{i-1}$ are independent, then the Chernoff bound is very useful, see for example Theorem 2.8 [38].

Theorem 7. *Let X be a random variable that can be expressed as a sum $X = \sum_{i=1}^{n} X_i$ of independent random indicator variables where $X_i \in \text{Bernoulli}(p_i)$ with (possibly) different $p_i = \mathbb{P}(X_i = 1) = \mathbb{E}[X_i]$. Then the following holds for $t \geq 0$:*

$$\begin{aligned} \mathbb{P}(X \geq \mathbb{E}[X] + t) &\leq \exp\left(-\frac{t^2}{2(\mathbb{E}[X] + t/3)}\right), \\ \mathbb{P}(X \leq \mathbb{E}[X] - t) &\leq \exp\left(-\frac{t^2}{2\mathbb{E}[X]}\right). \end{aligned}$$

In particular, if $\varepsilon \leq 3/2$, then

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon \mathbb{E}[X]) \leq 2\exp\left(-\frac{\varepsilon^2 \mathbb{E}[X]}{3}\right). \tag{3}$$

Strong concentration results of a sequence of not necessarily independent random variables can also be obtained if two consecutive random variables do not differ by too much and have the same (or similar) expectations. This is formalized using the concept of martingales, introduced next.

Definition 8. *A **martingale** is a sequence $X_0, X_1, \ldots$ of random variables defined on the random process such that*

$$\mathbb{E}[X_{n+1} \mid X_0, X_1, \ldots, X_n] = X_n.$$

In most applications, the martingale satisfies the property that

$$\mathbb{E}[X_{n+1} \mid X_0, X_1, \ldots, X_n] = \mathbb{E}[X_{n+1} \mid X_n] = X_n.$$

As a simple example, consider the following "random walk." Toss a coin n times. Let S_n be the difference between the number of heads and the number of tails after n tosses. S_n is a martingale. Indeed,

$$\mathbb{E}[S_{n+1} \mid S_n] = S_n + \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot (-1) = S_n.$$

Clearly, the expected value of S_n is zero. Thus, it is natural to expect that S_n stays relatively close to zero. The following well-known Hoeffding-Azuma inequality serves as a tool to investigate this, see for example Theorem 2.25 [38].

Lemma 9. *Let $X_0, X_1, \ldots$ be a martingale. Suppose that there exist constants $c_k > 0$ such that*

$$|X_k - X_{k-1}| \leq c_k$$

for each $k \leq n$. Then, for every $t > 0$,

$$\begin{aligned} \mathbb{P}(X_n \geq \mathbb{E}[X_n] + t) &\leq \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right), \\ \mathbb{P}(X_n \leq \mathbb{E}[X_n] - t) &\leq \exp\left(-\frac{t^2}{2\sum_{k=1}^n c_k^2}\right). \end{aligned}$$

This is often applied with t growing much faster than $\sqrt{n}$ and the c_k all small non-zero integers. In the martingale discussed above $c_k = 1$ for all k. Hence,

$$\mathbb{P}(|S_n| \geq \alpha\sqrt{n}) \leq 2\exp\left(\frac{(\alpha\sqrt{n})^2}{2n}\right) = 2\exp\left(\alpha^2/2\right),$$

which is arbitrarily small for α large enough.

Finally, let us mention that the Hoeffding-Azuma inequality can be generalized in many ways: an analogous inequality holds for supermartingales ($\mathbb{E}[X_{n+1}|X_n] \leq X_n$) as well as submartingales ($\mathbb{E}[X_{n+1}|X_n] \geq X_n$). Our proofs use the supermartingale method as described in Corollary 4.1 of [58]. We will use the following lemma, that follows from the generalized Hoeffding-Azuma inequality.

Lemma 10. *Let $G_0, G_1, \ldots, G_L$ be a random process and X_t a random variable determined by $G_0, G_1, \ldots, G_t$, $0 \leq t \leq L$. Suppose that for some real β and γ,*

$$\mathbb{E}(X_t - X_{t-1} \mid G_0, G_1, \ldots, G_{t-1}) < \beta$$

and

$$|X_t - X_{t-1} - \beta| \leq \gamma$$

for $1 \leq t \leq L$. Then for all $\varepsilon > 0$,

$$\mathbb{P}\big(\textit{For some } t \textit{ with } 0 \leq t \leq L : X_t - X_0 \geq t\beta + \varepsilon\big) \leq \exp\Big(-\frac{\varepsilon^2}{2L\gamma^2}\Big).$$

6.2 The role of the parameter ℓ

The parameter ℓ controls the relative speed of the influence and rewiring process in the network. Note that setting $\ell = 0.5$ (such that influence and rewiring occurs with equal probability at each time step) will cause attribute updating (via influence) to occur for a given node far more frequently than link rewiring (via homophily) for a given edge. This is a result of most networks having far more links than nodes.

To allow for the direct specification of the relative speed of rewiring versus influence, let $\ell = \ell(n, m)$ to be a function of the number of nodes and edges, and let ℓ tend to zero as $n \to \infty$. Let ℓ_0 be such that

$$\frac{\ell_0}{1-\ell_0} = \frac{n}{m},$$

and consider the evolutionary process run with parameter ℓ_0. Note that when $t = cn/\ell_0$ steps are performed (for some $c \in (0, \infty)$), we expect $\ell_0 t = cn$ influence steps and so an $1 - e^{-c}$ fraction of vertices should be hit. On the other hand, after t steps, the number of rewirings is expected to be $(1-\ell_0)t = cn(1-\ell_0)/\ell_0 = cm$ and so we expect to hit an $1 - e^{-c}$ fraction of edges as well. Note also that $\ell_0 = (1 + o(1))n/m = o(1)$, provided that $n = o(m)$ which is usually the case. For $n = cm$ for some $c \in (0, 1)$ we have $\ell_0 = c/(1+c) > 0$. There are three possible behaviors that might be desired for different scenarios:

- the two processes are of the same speed: $\ell = \ell_0$,
- there are more influence steps than rewiring ones: $\ell > \ell_0$,
- there are more rewiring steps than influence steps: $\ell < \ell_0$.

6.3 Proof of Lemma 1

Lemma 1 specifies the maximum length of time that is required to rewire each edge endpoint at least one time, and is stated in the main paper. In particular, Lemma 1 says that the time at which each edge endpoint has been rewired at least once is no later than $\hat{T}$, which is of order $m \ln m$.

Lemma. *Consider the evolutionary process $\mathfrak{P}(G_0, \omega_0, \ell, q, r, K) = (G_t, \omega_t)_{t=0}^{\infty}$. Let $\omega(m)$ be any function that tends to infinity arbitrarily slowly as $m \to \infty$, and let T be the first time every edge has both endpoints rewired at least once. Then, a.a.s.*

$$T \leq \frac{2Km}{1-\ell}\left(\ln m + \omega(m)\right) = \Theta(m \ln m). \tag{4}$$

Proof. Define $T^* = \frac{2Km}{1-\ell}\left(\ln m + \omega(m)\right)$ for some function $\omega(m)$ tending to infinity arbitrarily slowly. Fix a permutation of length $2m$ of all edge endpoints in the initial network G_0. For $i \in \{1, 2, \ldots, 2m\}$, let A_i be the event that the ith endpoint is *not* rewired after T^* steps. Note that at every step of the process, a given endpoint is rewired with probability at least $(1-\ell)/(2Km)$. Therefore, independently of the previous steps, we have that for every $i \in \{1, 2, \ldots, 2m\}$, and denoting by $\omega(m)$ any function tending to infinity with m arbitrarily slowly,

$$\begin{aligned}
\mathbb{P}(A_i) &\leq \left(1 - \frac{1-\ell}{2Km}\right)^{T^*} = \exp\left(-\frac{(1-\ell)T^*}{2Km} + O\left(\frac{T^*}{m^2}\right)\right) \\
&= (1 + o(1)) \exp\left(-\ln m - \omega(m)\right) \\
&= o(m^{-1}).
\end{aligned}$$

Therefore, the expected number of endpoints not rewired is $2m \cdot o(m^{-1}) = o(1)$ and so a.a.s. every endpoint is rewired at least once by Markov's inequality. $\diamond$ $\square$

6.4 Proof of Lemma 2

After time T, it is possible to make certain inferences about the distance of local centers of mass from the global center of mass. These results are specified in Lemma 2.

Lemma. *Consider the social process $\mathfrak{P}(G_0, \omega_0, \ell, q, 1, K) = (G_t, \omega_t)_{t=0}^{\infty}$. Suppose that $d = 2m/n$ is such that $\ln^7 n \ll d = o(n/\ln^3 n)$. Let $C > 0$. Then, w.e.p. for every $T \le t \le n^C$ and every $v \in V_t$, the following holds*

(i) If $K = 1$, then

$$M_t(v) = (1 + o(\ln^{-3} n))M_t.$$

(ii) If $K \ge 2$, then

$$M_t(v) \ge (1 + o(1))K^{-6}M_t + o(\ln^{-5} n)$$

and

$$M_t(v) \le 1 - (1 + o(1))K^{-6}(1 - M_t) + o(\ln^{-5} n).$$

Proof. As noted above, each dimension may be treated independently. Thus, without loss of generality we may assume that $r = 1$ and $\omega_t : V_t \to [0, 1]$. Fix $t \ge T$ and $v \in V_t$. Since we consider a polynomial number of events only, it is enough to show that each of them holds w.e.p.

The proof of part (i) is a simple application of the Chernoff bound. Let

$$X = \sum_{u \in N_t(v)} \omega_t(u) = |N_t(v)|M_t(v);$$

that is, X is the sum of distances from the beginning of the interval $[0, 1]$ to all the neighbors of v. As we already noticed, $M_t(v) = X/|N_t(v)|$, so it remains to estimate the random variable X and the degree of v independently. Note that, without loss of generality, we may assume that $M_t = M_t(n) \ge 1/2$, since there is no difference whether we measure the distance from the beginning of the interval $[0, 1]$ or from the end. For any fixed value of M_t, it follows from (2) that

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{u \in V_t} \mathbb{P}(vu \in E_t)\omega_t(u) \\
&= (1 + o(\ln^{-3} n))\frac{2m}{n^2} \sum_{u \in V_t} \omega_t(u) \\
&= (1 + o(\ln^{-3} n))\frac{2m}{n} M_t \\
&= (1 + o(\ln^{-3} n))dM_t.
\end{aligned}$$

Hence, $\mathbb{E}[X] = \Theta(m/n) = \Theta(d) \gg \ln^7 n$. It follows from (3), applied with $\varepsilon = d^{-3/7} = o(\ln^{-3} n)$, that w.e.p. $X = (1 + o(\ln^{-3} n))\mathbb{E}[X]$. Finally, using the Chernoff bound one more time to estimate the degree of v, we get that w.e.p. $|N_t(v)| = (1 + o(\ln^{-3} n))d$ and the proof of part (i) is finished.

In order to prove part (ii), it is convenient to think about the rewiring process as before (see the discussion right before (2)). Since each endpoint is rewired at least once at a given time $t \geq T$, we can fix an endpoint of some edge and then focus on the last *attempt* to rewire this endpoint. With probability at least $1/K$ (and clearly at most 1) the rewiring is performed and a given vertex is chosen with probability at least $1/(Kn)$ (at most K/n, respectively). This implies that for any endpoint and any vertex x with probability at least $1/(K^2n)$ (at most K/n, respectively) this endpoint is associated with x. Note that these bounds hold independently of the previous steps already performed. Performing the same calculations as in (2) we get

$$\mathbb{P}(xy \in E_t) \geq 1 - \left(1 - \frac{2}{K^4n^2}\right)^m = (1 + o(\ln^{-3} n))\frac{2m}{K^4n^2},$$

and

$$\mathbb{P}(xy \in E_t) \leq 1 - \left(1 - \frac{2K^2}{n^2}\right)^m = (1 + o(\ln^{-3} n))\frac{2K^2m}{n^2}.$$

Since $M_t(v) \geq 0$ for every v (deterministically), the statement trivially holds when $M_t = o(\ln^{-5} n)$. Hence, we may assume that $M_t = \Omega(\ln^{-5} n)$. Using the notation as in part (i), we have

$$\mathbb{E}[X] \geq (1 + o(1))\frac{2mM_t}{K^4n} = (1 + o(1))\frac{dM_t}{K^4} \gg \ln^2 n.$$

It follows from (3), applied with $\varepsilon = \ln^{-1/2} n$, that w.e.p. $X = (1 + o(1))\mathbb{E}[X]$. Similarly, we get that w.e.p. $|N_t(v)| \leq (1 + o(1))K^2d$ and so w.e.p. $M_t(v) \geq (1 + o(1))M_t/K^6$. The second statement of part (ii) is analogous, since there is no difference whether we measure the distance from the beginning of the interval $[0, 1]$ or from the end. The proof of part (ii) is complete. ◇ □

6.5 Explanation and proof of Theorem 3

To prepare for the slightly technical argument that is required to analyze the original model, a few simple processes are investigated first in the absence of homophily—that is, when $K = 1$ and edge rewirings are independent of vertex attributes. These results may then be applied with some modifications to the unconstrained case where $K \geq 1$.

We begin with a modified influence process (Model α) in which a vertex selected for attribute updating is moved to the global center of mass instead of the local center of mass. A second model (Model β) investigates a process in which there is no bias towards short edges ($K = 1$) and vertex attributes shift rapidly ($q = 1$) before generalizing the process (Model γ) to a slower version where attributes are more resistant to change ($q < 1$).

Since the first steps of the process are strongly influenced by the initial network and the initial vertex attributes, the beginning of the process is unpredictable. In the following analyses we wait T steps for every edge endpoint to be refreshed at least once (a.a.s. $T = \Theta(m \ln m)$ by Lemma 1). As noted above, each attribute dimension may be treated independently and so, without loss of generality, we assume that $r = 1$ and all attributes are in the interval $[0, 1]$. All of the following results may be generalized to any dimensionality of vertex attributes $r \in \mathbb{N}$.

Model α. For $t \geq T$, let X_t be the random variable equal to n times the global center of mass at time t; that is,

$$X_t = nM_t = \sum_{v \in V_t} \omega_t(v).$$

We may assume that $X_T = cn$ for some $c = c(n) \geq 1/2$, since there is no difference whether we measure the distance from the beginning of the interval $[0, 1]$ or from the end. At any time $t > T$ of the process, if an influence step is performed then we assume that a vertex is selected uniformly at random and is moved to the global center of mass (that is, to X_{t-1}/n). Network rewiring does not influence the value of X_t, so it is sufficient to concentrate on influence steps only. It follows from the well-known phenomenon of the coupon collector that a.a.s. after

$$\bar{T} = (n(\ln n + \omega(n)))/\ell = \Theta(\ell^{-1} n \ln n) \tag{5}$$

steps there are $n(\ln n + \omega(n) + o(1))$ influence steps and so a.a.s. every vertex's attribute has been updated at least once. At this point of the process every vertex is very close to the center of mass. In fact, we want to show that a.a.s. $\omega_{T+\bar{T}}(v) = c + o(\ln^{-2} n) = c + o(1)$ for every $v \in V_{T+\bar{T}}$.

It is easy to see that $\{X_t\}_{t \geq T}$ is a martingale. Indeed, by defining $L_v(t)$ as the event that vertex v was selected for influence at time t, we have

$$\begin{aligned} \mathbb{E}[X_t - X_{t-1} | X_{t-1}] &= \sum_{v \in V_{t-1}} \left(\frac{X_{t-1}}{n} - \omega_{t-1}(v) \right) \mathbb{P}(L_v(t)) \\ &= \left(X_{t-1} - \sum_{v \in V_{t-1}} \omega_{t-1}(v) \right) \frac{1}{n} = 0, \end{aligned}$$

provided that an influence step occurs at time t. Moreover, since only one vertex is affected at this step, $|X_t - X_{t-1}| \leq 1$. It follows from Hoeffding-Azuma's inequality (see Lemma 9) that w.e.p. during a period of $n(\ln n + \omega(n) + o(1))$ influence steps, X_t moves from the original value of X_T by at most $\omega(n)\sqrt{n} \ln n$. We get that w.e.p. the center of mass at any time t ($T \leq t \leq T + \bar{T}$) is at most

$$\frac{\omega(n)\sqrt{n} \ln n}{n} = \frac{\omega(n) \ln n}{\sqrt{n}} = o(\ln^{-2} n)$$

away from the starting point. The claim holds, since a.a.s. the attribute of every vertex has been updated in the meantime.

Model β. We now consider the process where homophily is not at work (that is, $K = 1$) and vertices selected for influence automatically adopt the average attribute of their neighbors ($q = 1$). The behavior of this model is very similar to the behavior observed in Model α. The only difference is that instead of moving to the global center of mass, vertices selected for influence processes are moved to their respective centers of mass. It follows from Lemma 2(i), however, that these local centers of mass are very close to the global center of mass.

In order to obtain a comparable ratio between influence and rewiring steps, we take $\ell = \Theta(\ell_0) = \Theta(n/m)$ (ℓ_0 is the threshold defined earlier) and assume that $\ell \geq n^{-C}$ for some $C \geq 1$ to cover a reasonable set of ℓ.[10] A result for this model is given in Theorem 11.

Theorem 11. *Consider the social process $\mathfrak{P}(G_0, \omega_0, \ell, 1, r, 1) = (G_t, \omega_t)_{t=0}^{\infty}$. Suppose that $d = 2m/n$ is such that $\ln^7 n \ll d = o(n/\ln^3 n)$ and $\ell \geq n^{-C}$ for some $C \geq 1$, and recall the definition of T in (1). Then, a.a.s.*

$$\|\omega_{T+\bar{T}}(v) - M_T\| = o(\ln^{-2} n) = o(1)$$

for every $v \in V_{T+\bar{T}}$, where $\bar{T}$ is defined in (5).

[10] As noted previously, for a comparable ratio of influence versus rewiring steps, we require $\ell = \Theta(n/m) = \Omega(n^{-1})$. In such a situation, we may take any $C > 1$. Larger values of C might be used to model processes in which influence steps are rare.

Note that a.a.s.

$$T + \bar{T} = \Theta(m \ln m) + \Theta(\ell^{-1} n \ln n) = O(n^{C+1} \ln n).$$

Proof. It follows from Lemma 2(i) that w.e.p. for every $T \le t \le T + \bar{T} \le 2m \ln m + 2\ell^{-1} n \ln n \le n^{3C}$ and every $v \in V_t$, we have $M_t(v) = (1 + o(\ln^{-3} n))M_t$. Hence, we may assume that this property holds deterministically in the time range we are interested in. The difference between the global center and the one of a vertex that is updating its attribute is at most $\beta = o(1/\ln^3 n)$.

As in the previous model, we try to investigate the behavior of the random variable $X_t = nM_t$ $(t \ge T)$ with $X_T = cn$. Without loss of generality, we may assume that $c = c(n) \ge 1/2$. This time, we need to use the generalized version of Hoeffding-Azuma inequality (see Lemma 10) that works for random variables that are not necessarily martingales. Since there are $O(n \ln n)$ influence processes, we can apply the lemma with $\beta = o(1/\ln^3 n), \gamma = 1.1, L = O(n \ln n), \varepsilon = \varepsilon(n) = \omega(n)\sqrt{n \ln n}$ (note that there is a lot of room for argument here; in fact, ε could be any function that is $o(n \ln^{-2} n)$) to get that a.a.s. for every $T \le t \le T + \bar{T}$, $X_t = (1 + o(\ln^{-2} n))X_T$ (note that $L\beta = o(n \ln^{-2} n)$). The claim holds, since a.a.s. every vertex was selected for influence in this time interval. $\diamond$ $\square$

Model γ. This model continues investigating a process without homophily ($K = 1$), but this time we relax the condition of rapid attribute updating through influence. We now consider cases where $q < 1$. In this case, the global center of mass still does not change much during the process. While in Model β it was enough for a vertex attribute to be updated at least once to arrive close to the global center of mass, this time the influence process is rather slow, and so each vertex must be selected many times in order to arrive very close to the global center of mass. The result is stated in the main paper, as Theorem 3.

To summarize, Theorem 3 tells us that, at time $T + \tilde{T}$ and for $K = 1$, the distance between each vertex's attribute and the global center of mass is of order less than $o(\ln^{-2} n)$. At this point, all vertex attributes are in some interval (a, b) of length $o(\ln^{-2} n)$ and, once all attributes are in this interval, they will always remain there.

Theorem. *Consider the evolutionary process $\mathfrak{P}(G_0, \omega_0, \ell, q, r, 1) = (G_t, \omega_t)_{t=0}^{\infty}$. Suppose that $d = 2m/n$ is such that $\ln^7 n \ll d = o(n/\ln^3 n)$ and $\ell \ge n^{-C}$ for some $C \ge 1$, and recall the definition of T in (1). Then, a.a.s.*

$$\|\omega_{T+\tilde{T}}(v) - M_T\| = o(\ln^{-2} n) = o(1)$$

for every $v \in V_{T+\tilde{T}}$, where $\tilde{T} = \ell^{-1}(2 + \varepsilon) n \ln n$ for some $\varepsilon > 0$.

Note that a.a.s.

$$T + \tilde{T} = \Theta(m \ln m) + \Theta(\ell^{-1} n \ln n) = O(n^{C+1} \ln n).$$

Proof. The proof is very similar to the proof of Theorem 11. We use Lemma 2(a) again to get that w.e.p. for every $T \le t \le T + \tilde{T}$ and every $v \in V_t$, we have $M_t(v) = (1 + o(\ln^{-3} n))M_t$.

We focus on the random variable $X_t = nM_t$ $(t \ge T)$ with $X_T = cn$. As usual, we may assume (without loss of generality) that $c = c(n) \ge 1/2$. We need to show first that $\{X_t\}_{t \ge T}$ is close to a martingale. Let

$L_v(t)$ be the event that v is selected for influence at time t. For any t, provided that an influence step occurs at that time,

$$\begin{aligned}
\mathbb{E}[X_t - X_{t-1}|X_{t-1}] &= \sum_{v\in V_{t-1}} \Big(\big(qM_{t-1}(v) + (1-q)\omega_{t-1}(v)\big) - \omega_{t-1}(v)\Big)\,\mathbb{P}(L_v(t)) \\
&= \frac{q}{n}\sum_{v\in V_{t-1}} \Big(M_{t-1}(v) - \omega_{t-1}(v)\big)\Big) \\
&= \frac{q}{n}\sum_{v\in V_{t-1}} \Big((1+o(\ln^{-3} n))M_{t-1} - \omega_{t-1}(v)\big)\Big) \\
&= \frac{q}{n}\Big((1+o(\ln^{-3} n))nM_{t-1} - \sum_{v\in V_{t-1}} \omega_{t-1}(v)\big)\Big) \\
&= \frac{q}{n}\Big((1+o(\ln^{-3} n))X_{t-1} - X_{t-1}\big)\Big) \\
&= o(\ln^{-3} n).
\end{aligned}$$

Arguing as before, we obtain that a.a.s. for every $T \le t \le T + \tilde{T}$,

$$X_t = (1+o(\ln^{-2} n))X_T$$

and so the global center of mass does not change much during this time.

Let us focus on some vertex v for now. Let $d_t = \|\omega_t(v) - c\|$ be the distance measured at time t between v and the global center of mass at time T; in particular, $d_T \le 1$. Note that every time a vertex v is selected during the influence step, the distance between this vertex and c shrinks by a constant factor. Formally, we know that

$$d_{t+1} = d_t(1-q) + o(\ln^{-2} n) = d_t(1-q+o(1)) \le d_t\left(1 - \frac{q}{2}\right),$$

provided that $d_t = \Omega(\ln^{-3} n)$. Let $\omega(n) = \Theta(\ln\ln n) = o(\ln n)$ be a function tending to infinity as $n \to \infty$ such that $(1-q/2)^{\omega(n)} = \ln^{-5/2} n$. Since after $\omega(n)$ influence steps $d_t = o(\ln^{-2} n)$, it remains to show that a.a.s. every vertex was selected to update its attributes $\omega(n)$ times in this time interval. Let $Y = Y(v)$ be a random variable counting how many times vertex v was selected. We have

$$\mathbb{E}[Y] = \frac{\tilde{T}\ell}{n} = (2+\varepsilon)\ln n.$$

It follows from the Chernoff bound (Theorem 7) that

$$\begin{aligned}
\mathbb{P}(Y \le \omega(n)) &= \mathbb{P}(Y \le \mathbb{E}[Y] - (\mathbb{E}[Y] - \omega(n))) \\
&\le \exp\left(-\frac{(\mathbb{E}[Y] - \omega(n))^2}{2\mathbb{E}[Y]}\right) \\
&= \exp\left(-\frac{(2+\varepsilon-o(1))^2}{2(2+\varepsilon)}\ln n\right) \\
&= \exp\left(-\left(1 + \frac{\varepsilon}{2} - o(1)\right)\ln n\right) \\
&= o(n^{-1}).
\end{aligned}$$

Hence, the expected number of vertices that have updated their attributes at most $\omega(n)$ times is $o(1)$. It follows from Markov's inequality that a.a.s. every vertex has updated its attributes at least $\omega(n)$ times. The proof is complete. $\diamond$ $\square$

6.6 Derivation and proof of Corollary 4

At time T, that is, when every endpoint is rewired at least once, vertex attributes might still be distributed across the whole space $[0,1]^r$. On the other hand, Theorem 3 implies that at time $T+\tilde{T}$ a.a.s. all vertices are in a ball of radius $o(\ln^{-2} n)$. Theorem 12, stated below, investigates the convergence process—that is, how the process behaves between time T and $T+\tilde{T}$.

For $T \le t \le T+\tilde{T}$, let Z_t/n be the average distance between vertices and the global center of mass at time T; that is,

$$Z_t = \sum_{v \in V_t} |\omega_t(v) - M_T|. \tag{6}$$

It is clear that $Z_T \le n$. (In fact, $Z_t \le n$ for all $t \ge 0$.) Suppose that at time t a vertex v is selected for influence. It follows from the proof of Theorem 3 that the global center of mass does not move much during the process. Moreover, from Lemma 2(i) we have that the local centers of mass are close to the global center of mass. Combining these two results, we get that the distance between the global center located at M_t and v decreases from its original value of $|\omega_{t-1}(v) - M_T|$ in a predictable way. Namely,

$$Z_t - Z_{t-1} = -|\omega_{t-1}(v) - M_T|q + o(\ln^{-2} n).$$

Hence,

$$\begin{aligned} \mathbb{E}[Z_t - Z_{t-1}|Z_{t-1}] &= \sum_{v \in V_{t-1}} \left(-|\omega_{t-1}(v) - M_T|q + o(\ln^{-2} n)\right)\ \mathbb{P}(L_v(t)) \\ &= -\frac{Z_{t-1}}{n} q\ell + o(\ln^{-2} n). \end{aligned}$$

This random variable is analyzed using the differential equation (DE) method [58]. Defining a real function $z(x)$ to model the behavior of Z_{T+xn}/n, the above relation implies the following differential equation

$$z'(x) = -q\ell z(x),$$

with the initial condition $z(0) = Z_T/n$. The general solution is

$$z(x) = \exp(-q\ell x + C),\ C \in \mathbb{R},$$

and the particular solution is $z(x) = (Z_t/n)\exp(-q\ell x)$. This *suggests* that the random variable Z_t should be close to the deterministic function

$$Z_T \exp(-q\ell(t-T)/n).$$

Theorem 12 below states precisely the conditions under which this holds.

Theorem 12. *Consider the social process $\mathfrak{P}(G_0, \omega_0, \ell, q, r, 1) = (G_t, \omega_t)_{t=0}^{\infty}$. Suppose that $d = 2m/n$ is such that $\ln^7 n \ll d = o(n/\ln^3 n)$ and $\ell \ge n^{-C}$ for some $C \ge 1$. Then, a.a.s.*

$$Z_t = (1+o(1)) Z_T \exp\left(-\frac{q\ell(t-T)}{n}\right)$$

for every t such that $T \le t \le T + t_f$, where T is defined in (1) and

$$t_f = \frac{n}{q\ell}\left(\ln\ln n - \ln\left(\frac{n}{Z_T}\right)\right).$$

Since $Z_{T+t_f} = n/\ln n$, this theorem tells us that, a.a.s., the random variable Z_t is well concentrated around the solution of the differential equation until it reaches the value of $n/\ln n$. Note that $t_f \le \tilde{T}$. It should be noted that Corollary 4, stated in the main paper, holds under this Theorem with Z_t replaced by W_t.

Before moving to a proof of Theorem 12, one more useful tool is needed: the "stopping time."

A **stopping time** with respect to a random process is a random variable S with values in $\{0, 1, \dots\} \cup \{\infty\}$ for which it can be determined whether $S = t$, for any time t, from knowledge of the process up to and including time t. The name can be misleading, since a process does not necessarily *stop* when it reaches a stopping time. The key result says that if a supermartingale (X_i) is stopped at a stopping time (that is, (X_i) becomes static for all time after the stopping time), then the result is a supermartingale; see, for example, [58] for more details.

Theorem 13. *If, with respect to some process, (X_i) is a supermartingale and S is a stopping time, then $(X_{\min\{i,S\}})$ is also a supermartingale with respect to the same process.*

Now, we are ready to prove Theorem 12:

Proof. Since the random variable Z_t does not change during rewiring steps, we focus on influence steps only. Suppose that the influence process occurs at times $T \le s_1 \le s_2 \le \ldots \le s_t \le \ldots$. Put $X_0 = Z_T$ and for $t > 0$, let $X_t = Z_{s_t}$. We need to transform X_t into something close to a martingale. Consider the following random variable

$$H_t = \ln X_t + \frac{qt}{n}$$

and the stopping time

$$S = \min\left\{t \ge 0 : X_t < \frac{n}{2\ln n} \vee t = t_f\ell\right\}.$$

(Note that H_t is chosen so that it is close to a constant along every trajectory of the differential equation $z'(x) = -qz(x)$.)

Consider the sequence of random variables $(H_t : 0 \le t \le t_f\ell)$. In order to use the generalized Hoeffding-Azuma inequality (see Lemma 10), we need to estimate the expected change:

$$\begin{aligned}
&\mathbb{E}[H_t - H_{t-1} \mid X_{t-1}, t-1] \\
&\quad= \sum_{v \in V_{s_{t-1}}} \left(\ln\left(Z_{s_{t-1}} - |\omega_{s_{t-1}}(v) - M_T|q + o(\ln^{-2} n)\right) - \ln Z_{s_{t-1}} + \frac{q}{n}\right) \\
&\qquad\cdot\ \mathbb{P}\left(L_v(s_t) \,\middle|\, \bigcup_{v \in V_{s_{t-1}}} L_v(s_t)\right) \\
&\quad= \frac{q}{n} + \frac{1}{n}\sum_{v \in V_{s_{t-1}}} \ln\left(1 - \frac{|\omega_{s_{t-1}}(v) - M_T|q + o(\ln^{-2} n)}{Z_{s_{t-1}}}\right).
\end{aligned}$$

By Taylor expansion,

$$\begin{aligned}
&\mathbb{E}[H_t - H_{t-1} \mid X_{t-1}, t-1] \\
&\quad = \frac{q}{n} + \frac{1}{n} \sum_{v \in V_{s_{t-1}}} \left(-\frac{|\omega_{s_{t-1}}(v) - M_T|q + o(\ln^{-2} n)}{Z_{s_{t-1}}} + O\left(\frac{1}{Z^2_{s_{t-1}}}\right)\right) \\
&\quad = \frac{q}{n} + \frac{1}{n}\left(-\frac{\sum_{v \in V_{s_{t-1}}} |\omega_{s_{t-1}}(v) - M_T|q}{Z_{s_{t-1}}}\right) + o\left(\frac{1}{Z_{s_{t-1}} \ln^2 n}\right) \\
&\quad = o\left(\frac{1}{X_{t-1} \ln^2 n}\right) = o\left(\frac{1}{n \ln n}\right),
\end{aligned}$$

provided $S > t$, since then $X_{t-1} \geq \frac{n}{2 \ln n}$. Similarly, we get that

$$\begin{aligned}
|H_t - H_{t-1}| &\leq |\ln(X_{t-1} - 1) - \ln X_{t-1}| + \frac{q}{n} \\
&= O\left(\frac{1}{X_{t-1}}\right) + \frac{q}{n} = O\left(\frac{\ln n}{n}\right),
\end{aligned}$$

provided $S > t$ again.

Therefore, with $i \wedge S$ denoting $\min\{i, S\}$, we have

$$\begin{aligned}
\mathbb{E}[H_{t\wedge S} - H_{(t-1)\wedge S} \mid X_{t-1}, t-1] &= o\left(\frac{1}{n \ln n}\right) \text{ and} \\
|H_{t\wedge S} - H_{(t-1)\wedge S}| &= O\left(\frac{\ln n}{n}\right).
\end{aligned}$$

Now we may apply Lemma 10 to the sequence $(H_{t\wedge S} : 0 \leq t \leq t_f \ell)$, and symmetrically to $(-H_{t\wedge S} : 0 \leq t \leq t_f \ell)$, with $\varepsilon = 1/\ln n$, $\beta = o(1/(n \ln n))$, and $\gamma = O(\ln n/n)$ to show that a.a.s. for all t with $0 \leq t \leq t_f \ell$ we have

$$|H_{t\wedge S} - H_0| = o(1).$$

(Note that there are $L \leq t_f \ell \leq \tilde{T}\ell = O(n \ln n)$ steps and so $L\beta = o(1)$.) As $H_0 = \ln X_0$, it follows from the definition of H_t, that a.a.s. $\ln(X_t/X_0) = -qt/n + o(1)$ for $0 \leq t \leq S$, and thus

$$X_t = (1 + o(1))X_0 \exp\left(-\frac{qt}{n}\right). \tag{7}$$

Now, we will show that a.a.s. $S = t_f \ell$. The events asserted by the equation hold a.a.s. up until time S, as shown above. Thus, in particular, a.a.s.

$$\begin{aligned}
X_S &= (1 + o(1))X_0 \exp\left(-\frac{qS}{n}\right) \\
&\geq (1 + o(1))Z_T \exp\left(-\frac{qt_f \ell}{n}\right) \\
&= (1 + o(1))\frac{n}{\ln n}
\end{aligned}$$

which implies that $S = t_f$ a.a.s., and so a.a.s. (7) holds for $0 \le t \le t_f \ell$.

To complete the proof, it is enough to show that the concentration for the random variable X_t implies the concentration for the original random variable Z_t. Fix t such that $T \le t \le T + t_f$. It follows easily from the Chernoff bound (3) that w.e.p. from time T until time t there are $(t-T)\ell + O(n^{2/3})$ time steps at which influence processes are performed. As a result w.e.p.

$$\begin{aligned} Z_t &= X_{(t-T)\ell + O(n^{2/3})} \\ &= (1+o(1))X_0 \exp\left(-\frac{q(t-T)\ell}{n} + o(1)\right) \\ &= (1+o(1))Z_T \exp\left(-\frac{q(t-T)\ell}{n}\right), \end{aligned}$$

and the proof is finished. ◇ □

6.7 Proof of Theorem 5

Theorem. *Consider the evolutionary process $\mathfrak{P}(G_0, \omega_0, \ell, q, r, K) = (G_t, \omega_t)_{t=0}^{\infty}$. Suppose that $d = 2m/n$ is such that $\ln^7 n \ll d = o(n/\ln^3 n)$ and $\ell \ge n^{-C}$ for some $C \ge 1$. Then, a.a.s. for every pair $u, v \in V_t$ we have*

$$\begin{aligned} |\omega_t(u) - \omega_t(v)| &\le \sqrt{r}\left(1 - \frac{q}{13K^{12} n \ln n}\right)^{\lfloor \ell(t-T)/(3n\ln n)\rfloor} \\ &\le \sqrt{r}\exp\left(-\frac{q\lfloor \ell(t-T)/(3n\ln n)\rfloor}{13K^{12} n \ln n}\right) \end{aligned}$$

for every t such that $T \le t = O(\ell^{-1} n^2 \ln^2 n)$ (the random variable T is defined in (1)). In particular, if $t = cq^{-1}\ell^{-1} n^2 \ln^2 n$ for some constant $c > 0$, then a.a.s. for every pair $u, v \in V_t$ we have

$$|\omega_t(u) - \omega_t(v)| \le (1+o(1))\sqrt{r}\exp\left(-\frac{c}{39K^{12}}\right).$$

Proof. As always, we assume that $r = 1$ and that at some time $t_0 \ge T$ all vertices lie in the interval $[0,1]$. (Note that if the maximum distance in each dimension is at most x, then the maximum distance is at most $\sqrt{r}x$.) Let $\varepsilon = (12K^{12} n \ln n)^{-1}$. The proof strategy is to show that during at most $3\ell^{-1} n \ln n$ steps the interval shrinks by a factor of at least $(1+o(1))q\varepsilon$ (that is, the length of the interval will be at most $1 - (1+o(1))q\varepsilon$). Since the time intervals can be treated independently, the result follows.

Let us consider the following two cases.

Case 1: for every $t_0 \le t \le t_2 := t_0 + 1.5\ell^{-1} n \ln n$, we have $M_t \ge K^6 \varepsilon$. It follows from Lemma 2(ii) that w.e.p. for every t in the range under consideration and $v \in V_t$,

$$M_t(v) \ge (1+o(1))K^{-6} M_t \ge (1+o(1))\varepsilon.$$

Hence, all vertices selected (during the influence process) between time t_0 and t_2 are moved away from 0; in fact, the distance from 0 is at least $(1+o(1))q\varepsilon$. Since during this time interval there are $(1.5+o(1))n\ln n$ influence steps w.e.p., all vertices are selected at least once w.e.p. After that, all vertices are in the interval $[(1+o(1))q\varepsilon, 1]$ and the result holds.

Case 2: for some t_1 such that $t_0 \le t_1 \le t_2 := t_0 + 1.5\ell^{-1} n \ln n$, we have $M_{t_1} < K^6\varepsilon$. Our goal is to show that the global center of mass cannot drift too fast to the other end of the interval during another $1.5\ell^{-1} n \ln n$ steps. Since $M_{t_1} < K^6\varepsilon$, at time t_1 there are at most $2K^6\varepsilon n$ vertices at distance at most $1/2$ from 1, the other end of the interval. We will call these vertices *red*; the remaining ones are *blue*. We will couple the original process with the one in which all red vertices are moved (at time t_1) to $1/2$. This is the only difference between the two processes; everything else remains the same, including the rewiring process on the (random) graph G_t. Let $\omega'_t(v)$ denote the position of vertex v at time t in the coupled process; similarly, $M'_t(v)$ is the corresponding local center of mass of v at time t. We have $\omega_{t_1}(v) = \omega'_{t_1}(v)$ for all blue vertices; and $\omega'_{t_1}(v) = 1/2 \le \omega_{t_1}(v)$ for all red ones.

We consider both processes starting from time t_1. First, we will show (by induction on i) that after $i < 1.6n \log n$ influence steps we have the following two properties w.e.p.:

i) for every blue vertex we have $|\omega_t(v) - \omega'_t(v)| \le 2K^{12}\varepsilon i$,

ii) for every red vertex we have $|\omega_t(v) - \omega'_t(v)| \le 1/2$.

The property trivially holds for $i = 0$. Suppose that the property holds for some i and vertex v is selected for the influence process at time t. All remaining vertices do not move (in both processes) and so one needs to show the desired property for v only. It follows that w.e.p.

$$\begin{aligned}
\left|M_{t-1}(v) - M'_{t-1}(v)\right| &\le \frac{1}{|N_{t-1}(v)|} \sum_{u \in N_{t-1}(v)} \left|\omega_{t-1}(u) - \omega'_{t-1}(u)\right| \\
&\le \frac{1}{|N_{t-1}(v)|} \Big((2K^{12}\varepsilon i)(\#\text{ of blue neighbors}) \\
&\quad + (1/2)(\#\text{ of red neighbors}) \Big) \\
&\le \frac{1}{|N_{t-1}(v)|} \left((2K^{12}\varepsilon i |N_{t-1}(v)|) + (1/2)(2K^{12}\varepsilon |N_{t-1}(v)|) \right) \\
&\le 2K^{12}\varepsilon(i+1),
\end{aligned}$$

since w.e.p. there are at most $2K^{12}\varepsilon |N_{t-1}(v)|$ red neighbors of v. Indeed, by the upper and lower bounds for the probability of an edge, there are at most $4K^8 m\varepsilon/n$ red neighbors, and the size of the neighborhood in total is at least $d/K^4 = (2m)/(nK^4)$. Hence the fraction of red neighbors is at most $2K^{12}\varepsilon$. It follows that w.e.p. for each blue vertex we have

$$\begin{aligned}
|\omega_t(v) - \omega'_t(v)| &\le q|M_{t-1}(v) - M'_{t-1}(v)| + (1-q)|\omega_{t-1}(v) - \omega'_{t-1}(v)| \\
&\le q2K^{12}\varepsilon(i+1) + (1-q)2K^{12}\varepsilon i \le 2K^{12}\varepsilon(i+1).
\end{aligned}$$

Similarly, w.e.p. for each red vertex we have

$$\begin{aligned}
|\omega_t(v) - \omega'_t(v)| &\le q|M_{t-1}(v) - M'_{t-1}(v)| + (1-q)|\omega_{t-1}(v) - \omega'_{t-1}(v)| \\
&\le q2K^{12}\varepsilon(i+1) + (1-q)(1/2) \le 1/2,
\end{aligned}$$

where the last inequality follows by our assumption on i. The claim holds by induction.

We consider $1.5\ell^{-1} n \ln n$ steps of the process which selects all vertices w.e.p. Clearly, in the coupled process the global center of mass is always in the interval $[0, 1/2]$ (deterministically). Since $(2K^{12}\varepsilon)(1.5n \ln n) <$

1/3, w.e.p. blue vertices in the original process are close to their counterparts in the coupled one: Indeed, since the difference of the global center of mass coming from the contribution of red vertices is at most $2K^6\varepsilon$, and the contribution coming from blue vertices at most $1/3$, the global center of mass in the original process must be smaller than $5/6+2K^6\varepsilon$ w.e.p. But this implies that once all vertices are selected for the influence process, all of them must be far away from 1: By Lemma 2(ii) we then have that $M_t(v) \leq 1-(1+o(1))\frac{1}{6}K^{-6}+o(1)$. Hence, the distance of $M_t(v)$ from the right border of the interval is at least $\frac{1}{6}K^{-6}+o(1)$, and once the vertex undergoes the influence process, its distance from the right border is thus at least $q\frac{1}{6}K^{-6}+o(1)$. The proof is finished. $\diamond$ $\square$